

Retail-Giant Sales Forecasting Case Study

Group Name: Datalogists

1. Indira Kumari Saraswatula
2. Samadunnisha Sheik
3. Prudvi Bilakanti
4. Sampath Kanduri

“Global Mart” is an online store super giant having worldwide operations. It takes orders and delivers across the globe and deals with all the major product categories - consumer, corporate & home office.

Objective:

Forecast the sales and the demand for the next 6 months, for Global Mart to manage revenue and inventory accordingly.

Conditions/Constraints :

- The store caters to 7 different market segments and in 3 major categories.
- Need to subset the data into 21 (7×3) buckets to forecast the data at granular level
- Out of 21 buckets only 2 most profitable segment will be consider for forecasting.

Facts & Specifics about data:

The data currently has the transaction level data, where each row represents a particular order made on the online store.

There are 24 attributes related to each such transaction.

The “Market” attribute has 7-factor levels representing the geographical market sector that the customer belongs to. Eg: US, APAC, EU

The “Segment” attribute tells which of the 3 segments that customer belongs to. (Consumer, Home Office, Corporate)

Data Collection & Data Understanding

4- year of data to be subset into 21 (7*3) buckets

Data Cleaning

Check for NAs

Check for duplicates

Data Preparation

subset of the data frame on which we will be working further

Convert the Date attribute and Month attribute to the dataset

Convert the transaction-level data into a time series

Aggregate the 3 attributes - Sales, Quantity & Profit, over the Order Date at Mont level

Based on coefficient of variation of the Profit find the 2 profitable segments.

Model Building

Smoothen the curve using Moving Average Smoothing

Model Seasonality using sinusoid function

Check Residue using auto ARIMA

Evaluate Model

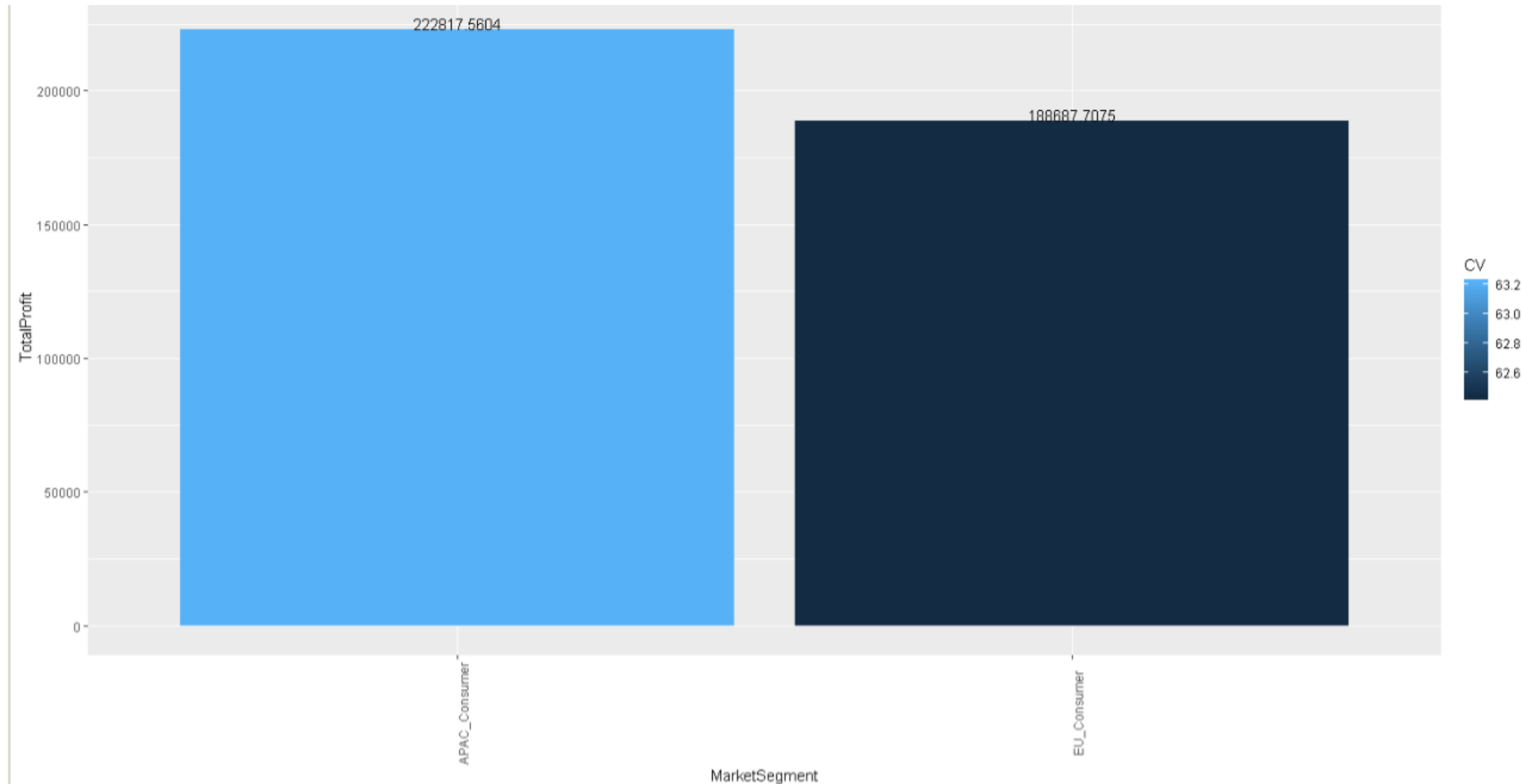
Check the residual series is white noise

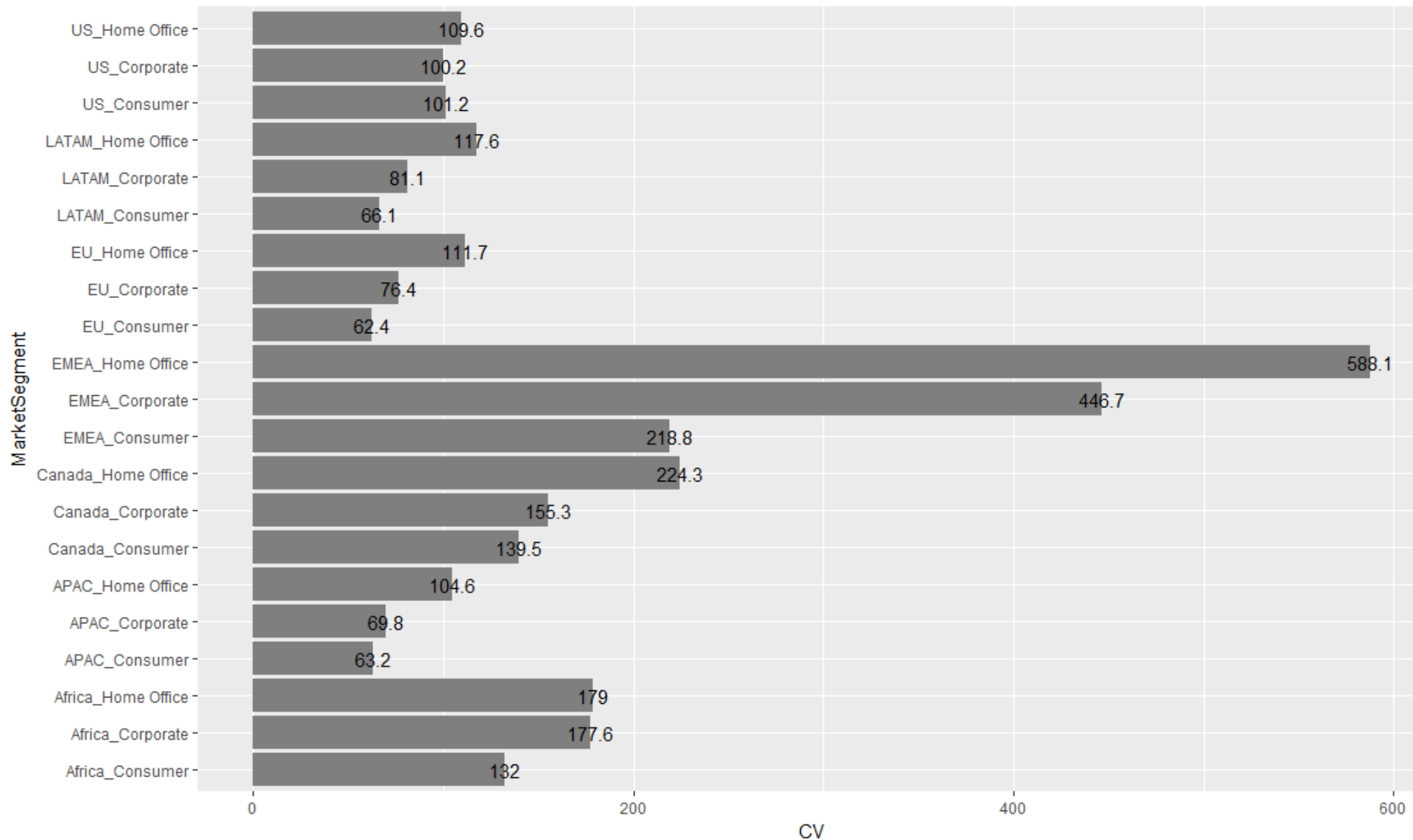
Evaluate the model using MAPE

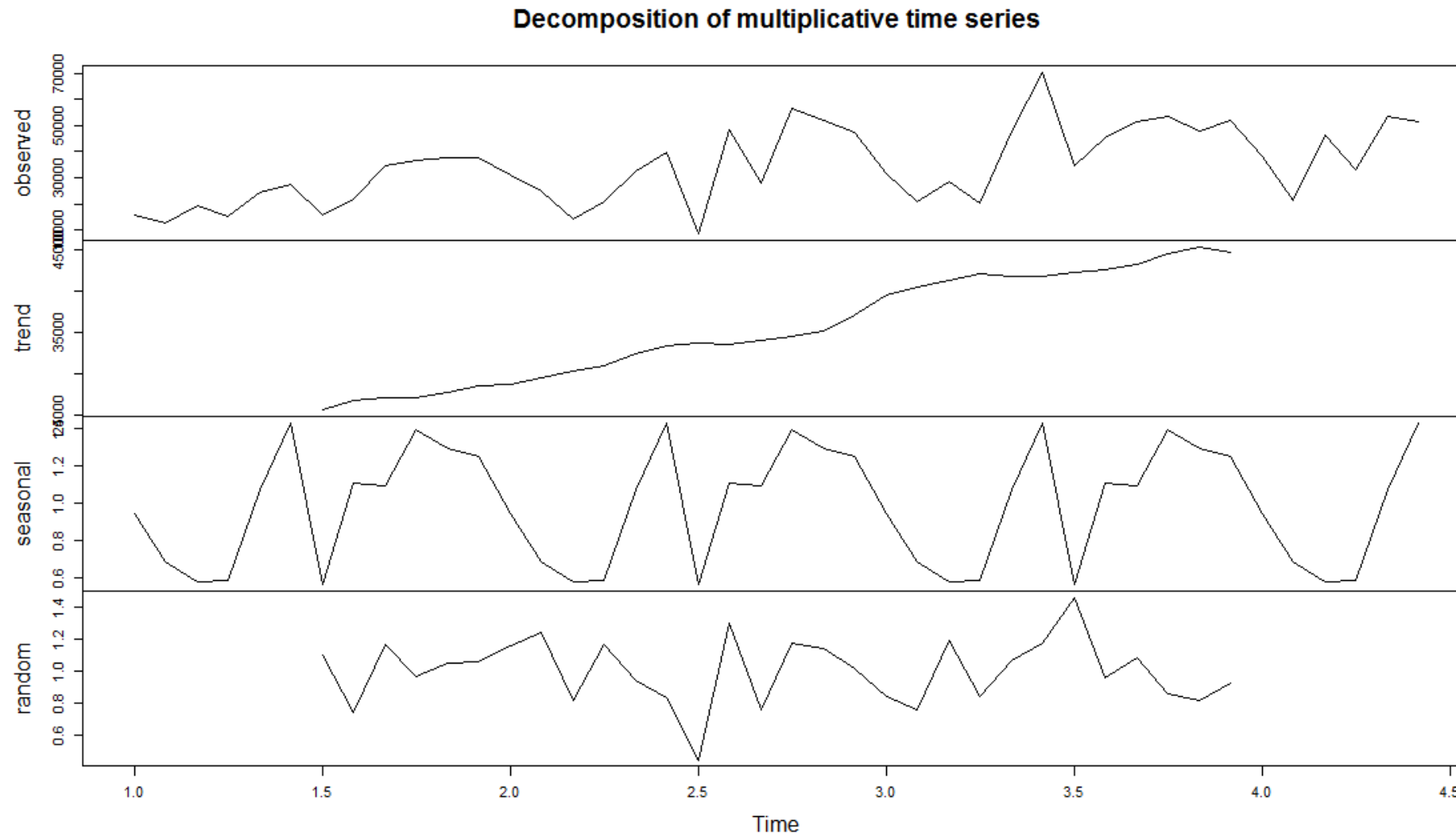
Model Accepted

forecast the sales/demand for next 6 months

Top 2 Significantly profitable Segments



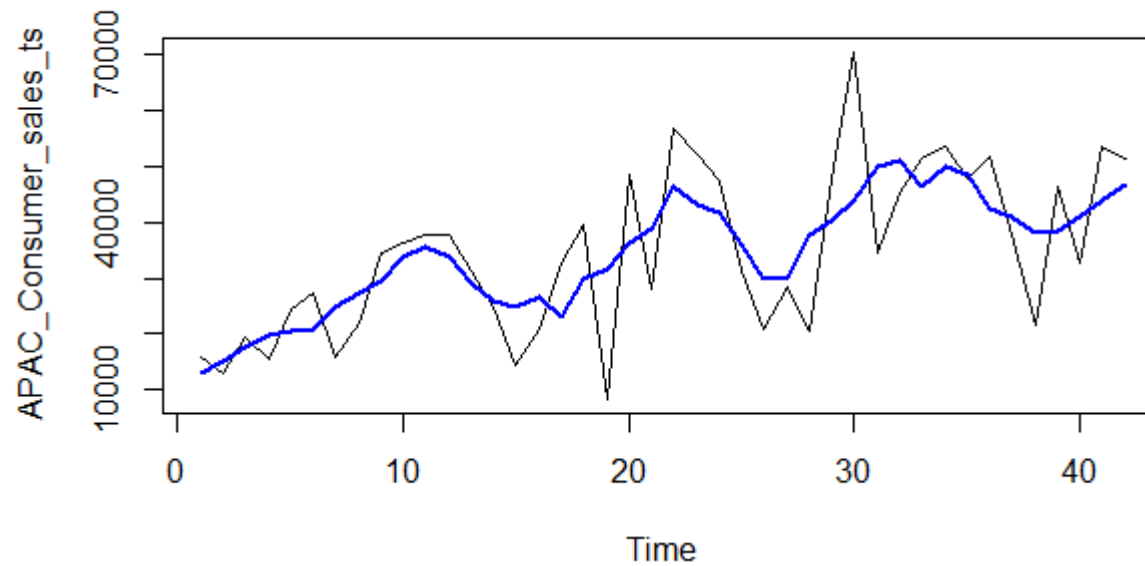




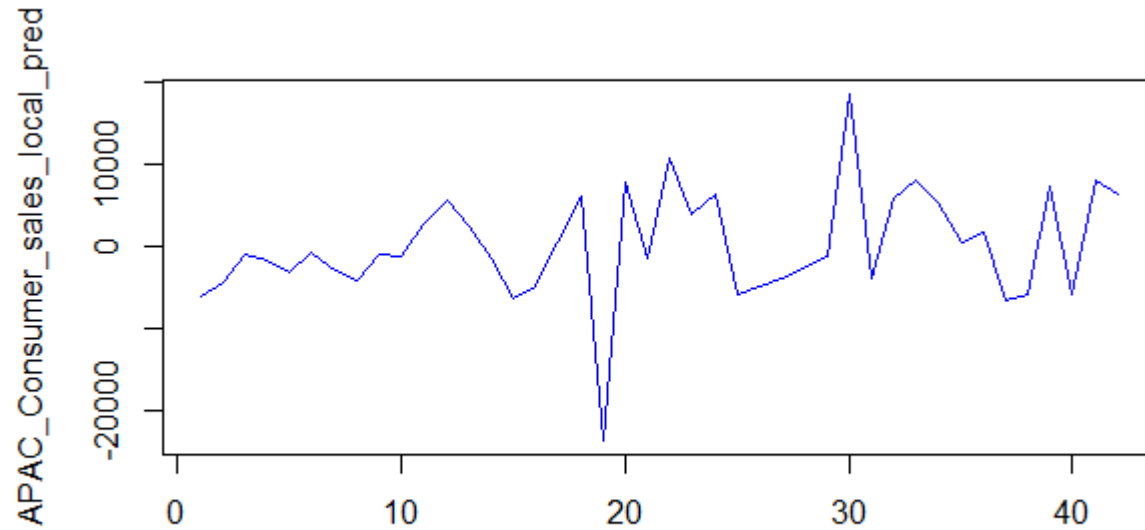
APAC Consumer – sales

Smoothing techniques are required to Isolate trend and seasonality

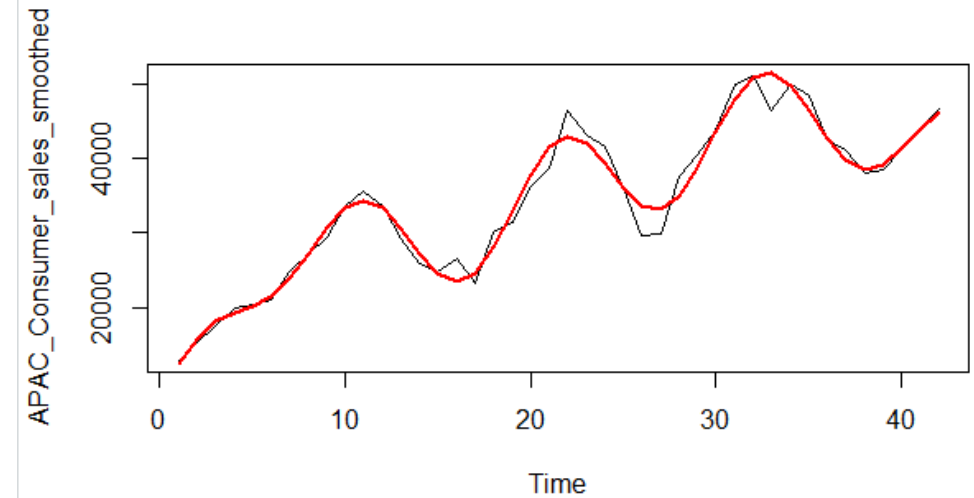
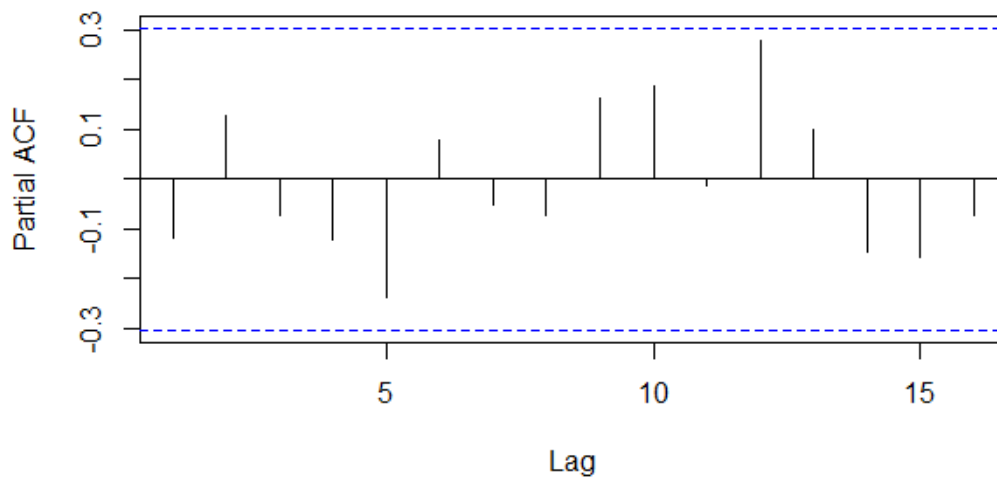
Blue line – shows the smooth time series



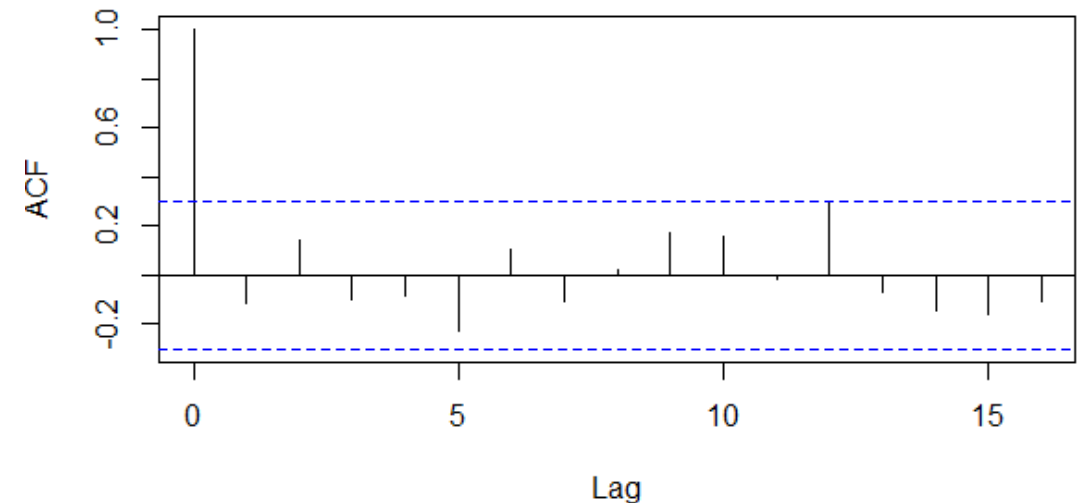
Used the 3rd degree polynomial by hit and trial



Series APAC_Consumer_sales_local_pred



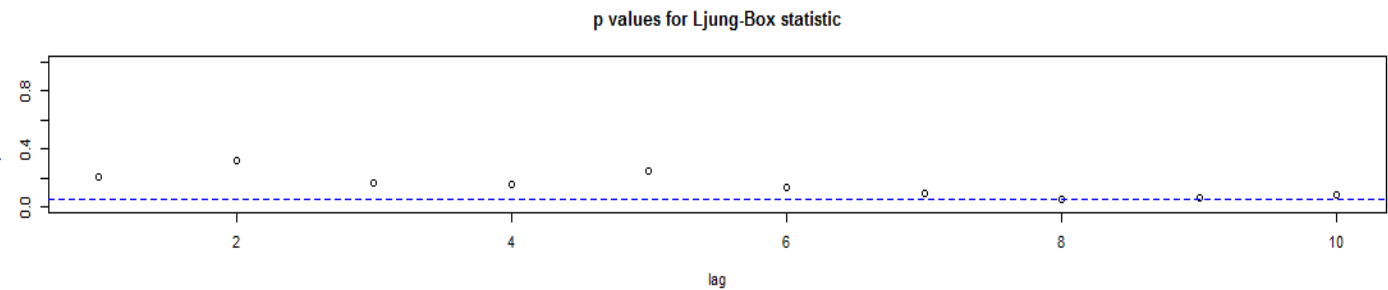
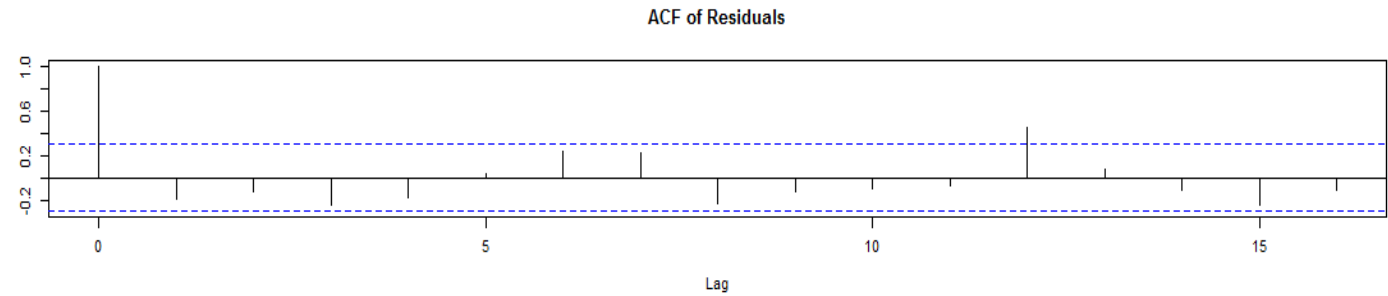
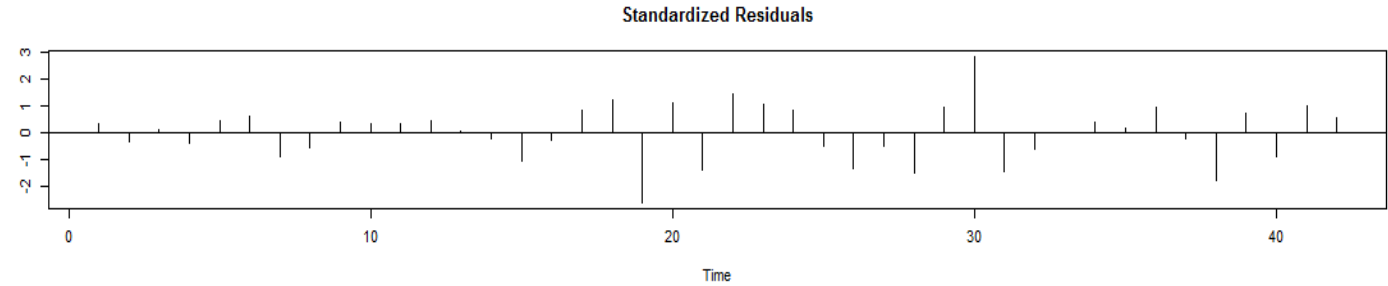
Series APAC_Consumer_sales_local_pred



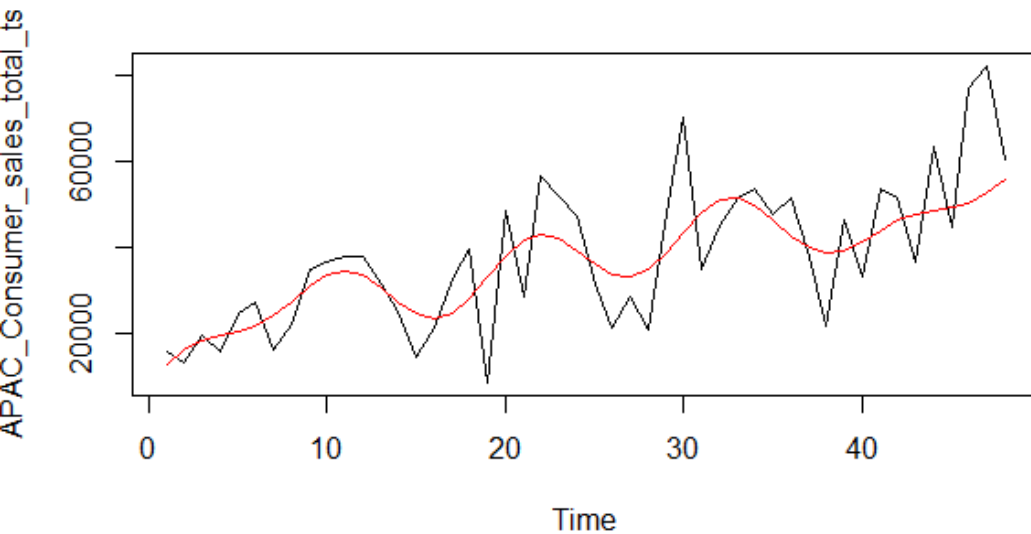
Model Diagnostics for a Fitted ARIMA Model

Series: APAC_Consumer_sales_local_pred
ARIMA(0,0,0) with zero mean

sigma² estimated as 91944190: log likelihood=-444.67
AIC=891.33 AICC=891.43 BIC=893.07



The Actuals Vs predictions

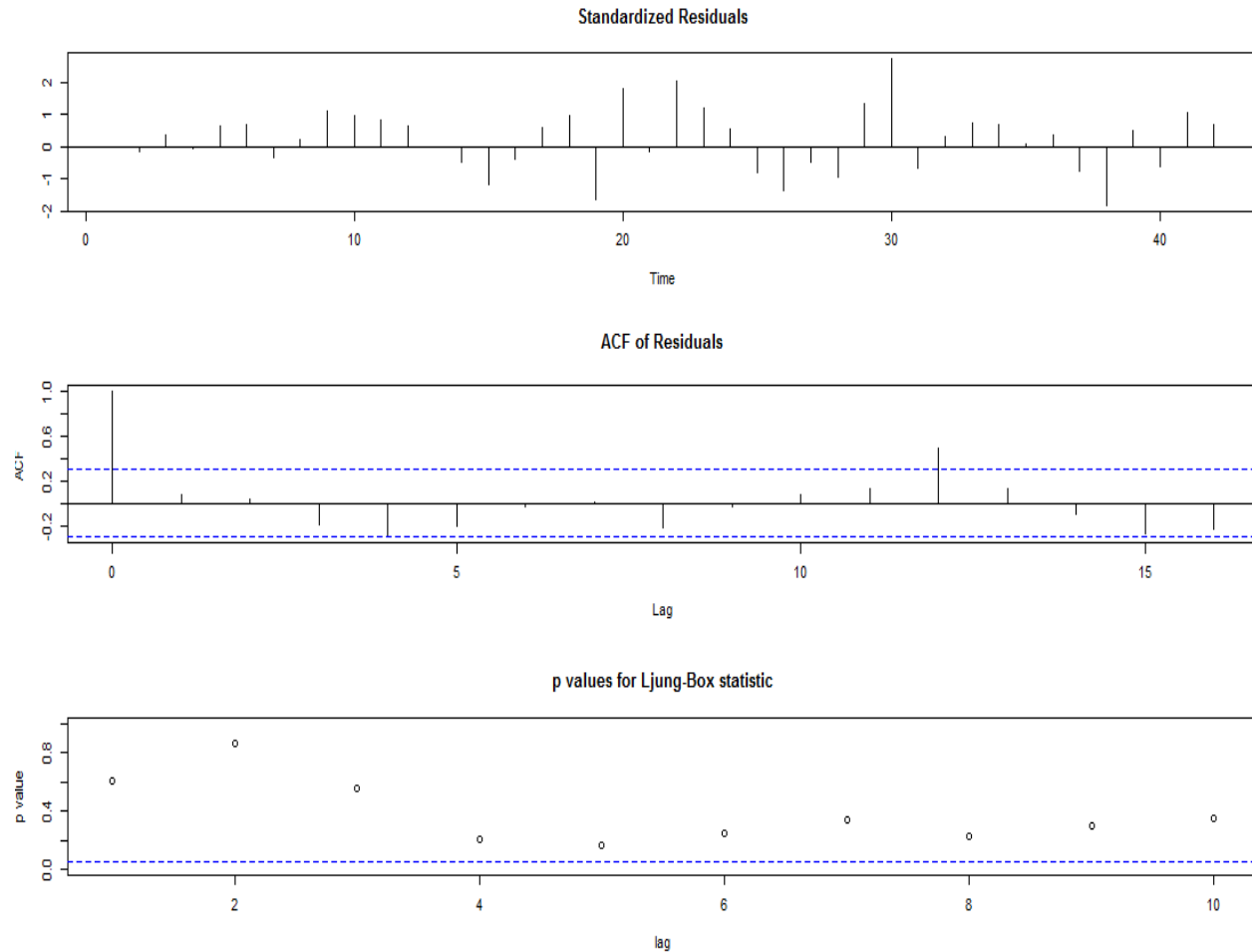
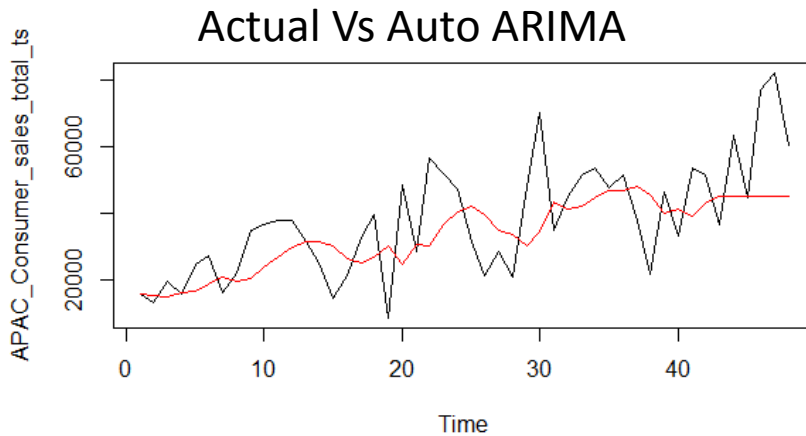


Model Diagnostics for a Fitted ARIMA Model

Series: APAC_Consumer_sales_ts
ARIMA(0,1,1)

Coefficients:
ma1
-0.7559
s.e. 0.1381

sigma² estimated as 174361555: log likelihood=-447.11
AIC=898.23 AICC=898.55 BIC=901.66



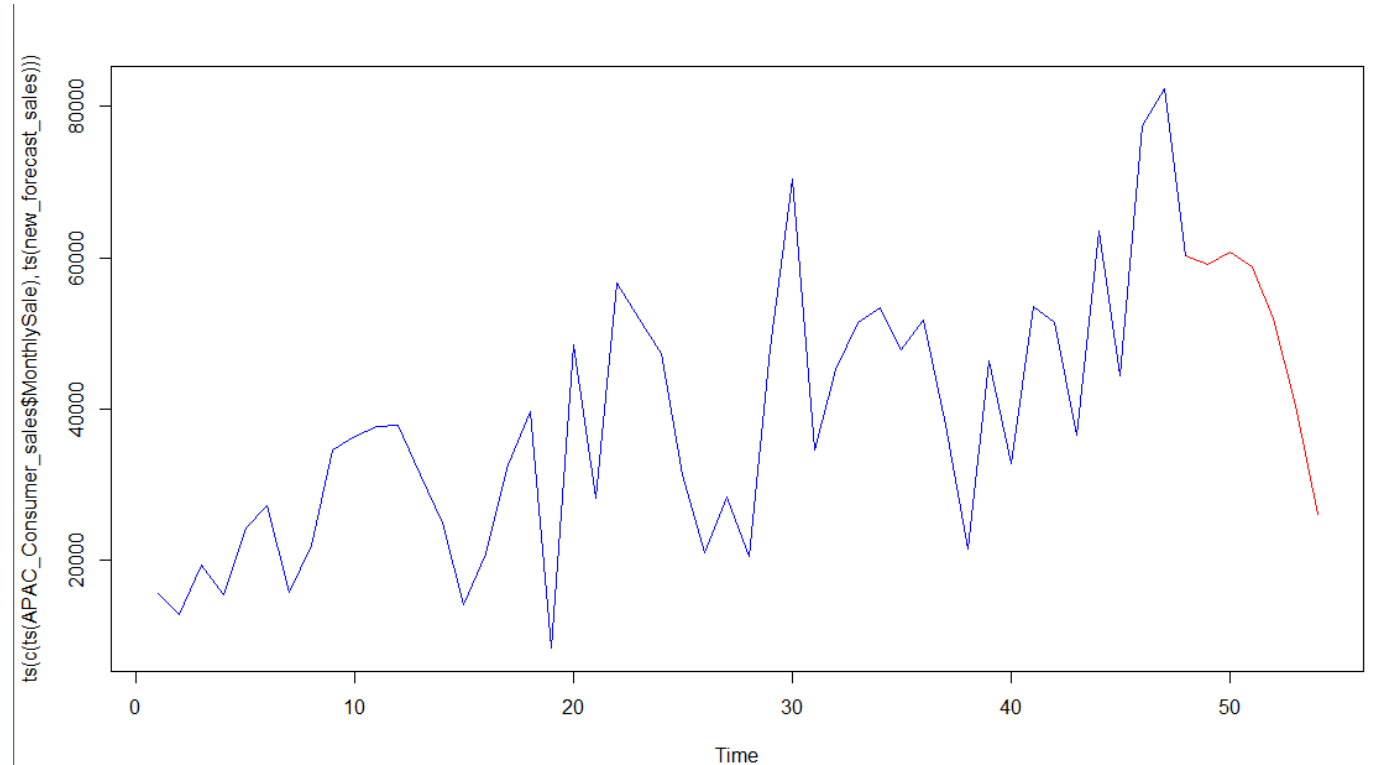
Conclusion:

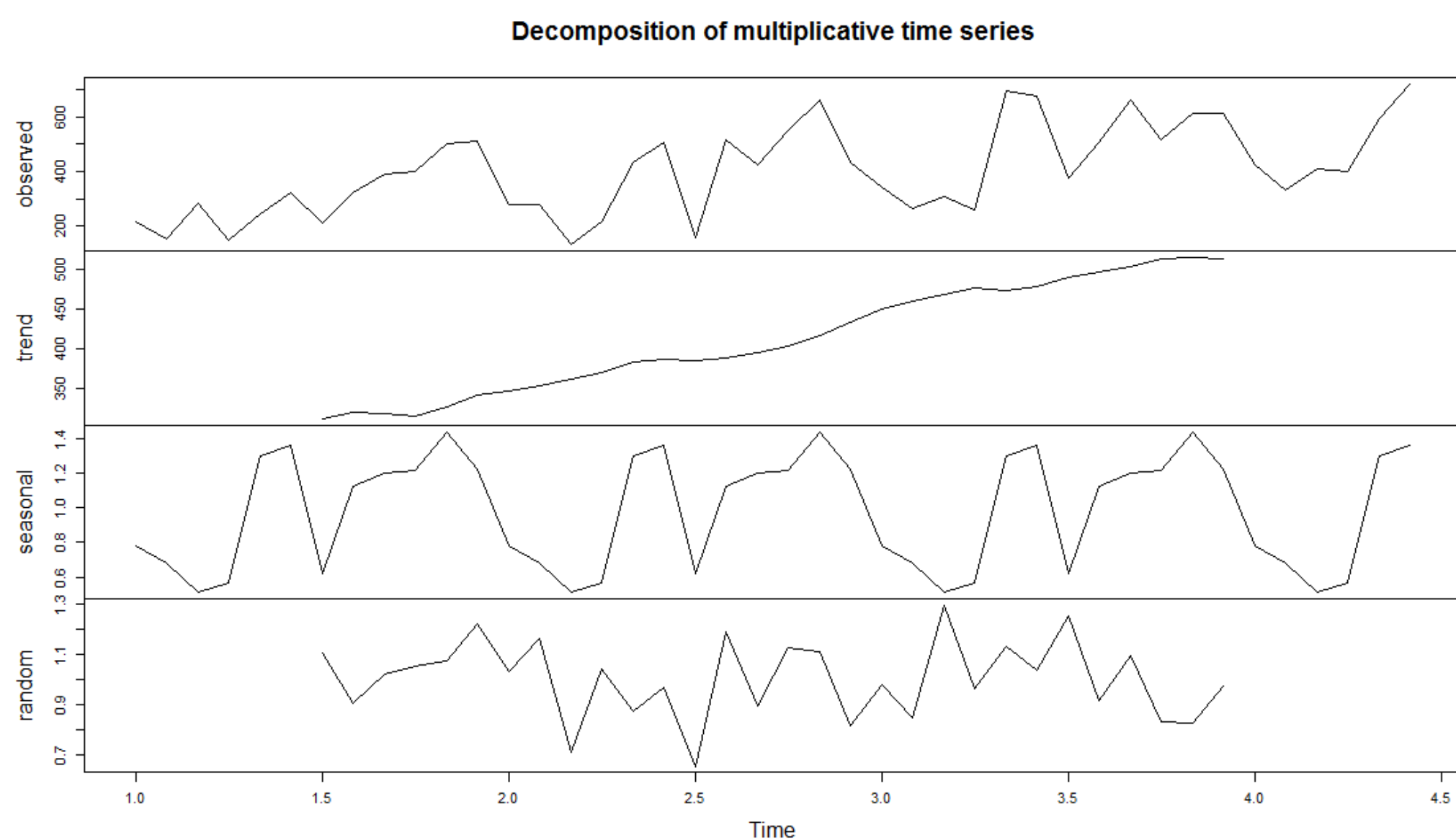
MAPE value of classical decomposition is less than that of AUTO ARIMA model. The log likelihood is also higher for classical decomposition. So classical decomposition is best model than ARIMA

As classical decomposition is doing better than ARIMA ,Forecast values for APAC Consumer Sales for next 6 Months using classical decomposition method

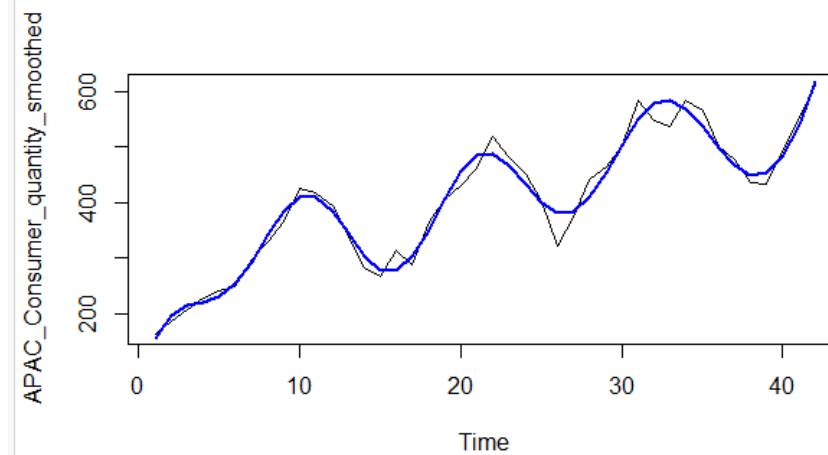
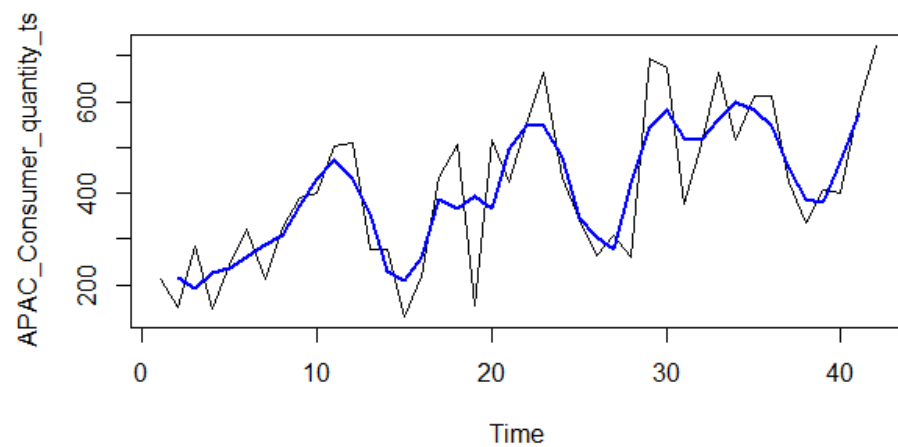
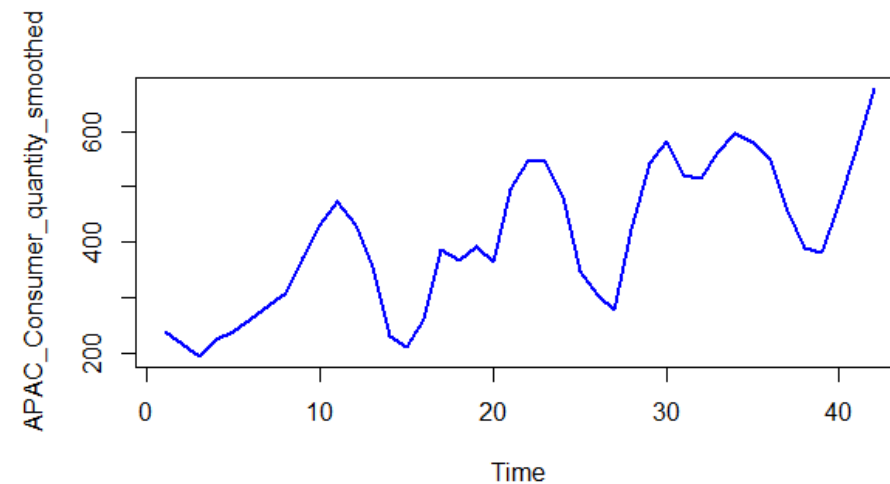
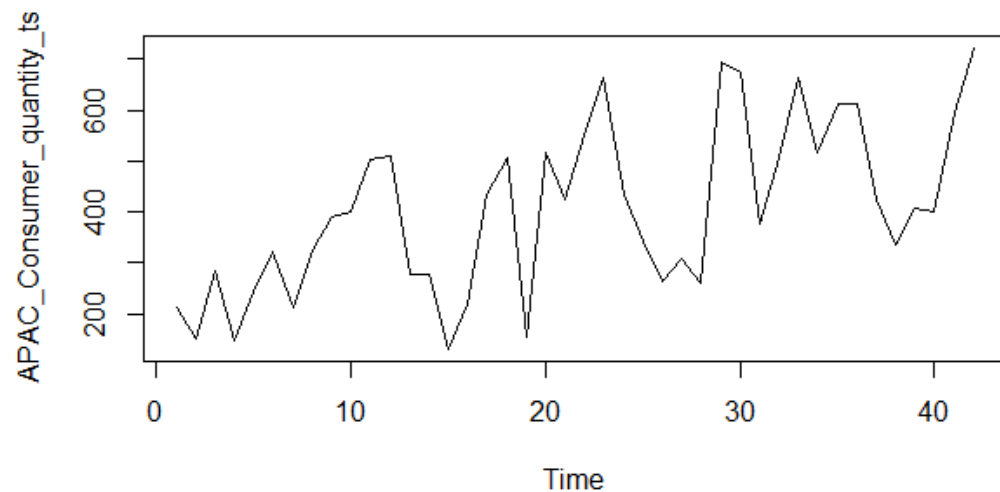
Forecast values for next 6 Months

1	2	3	4	5	6
59201.08	60790.11	58805.39	51988.28	40457.60	26152.28

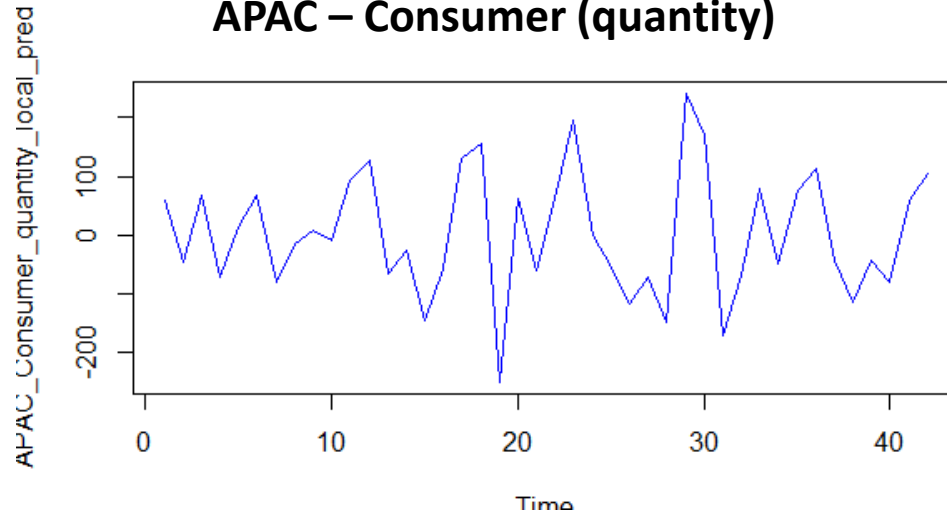




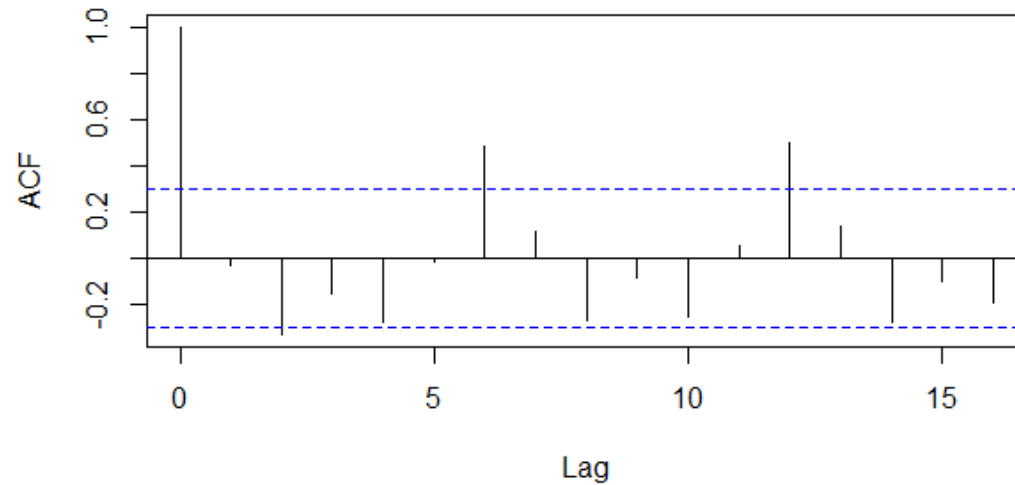
Smoothed Graph :: APAC – Consumer (quantity)



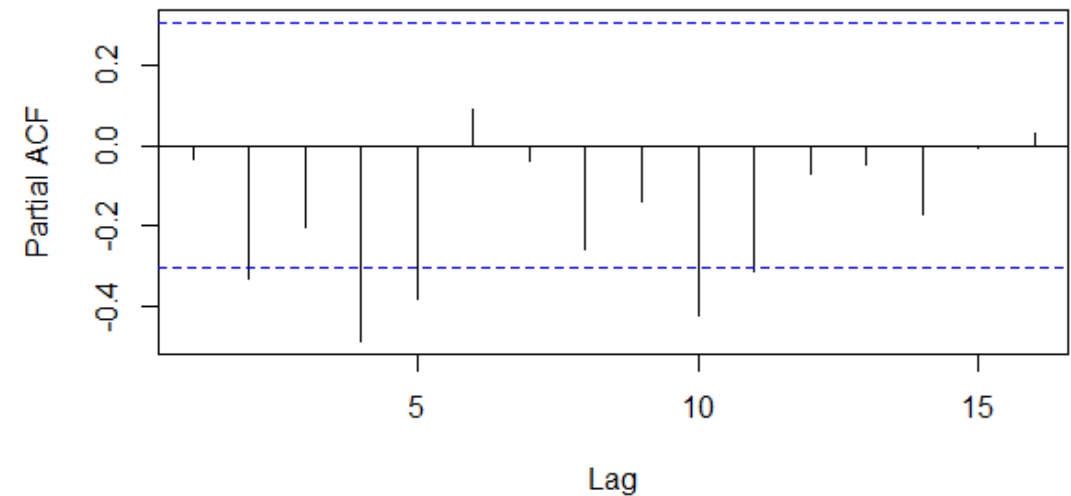
APAC – Consumer (quantity)



Series APAC_Consumer_quantity_local_pred



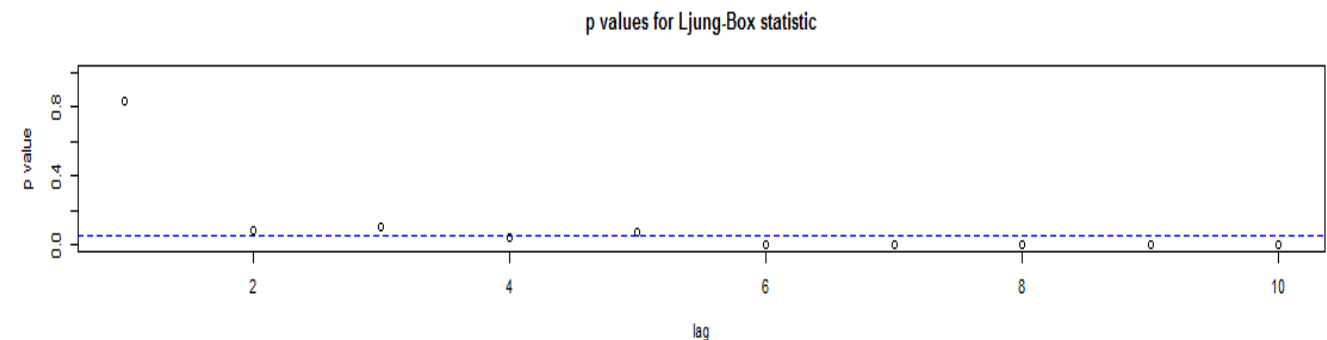
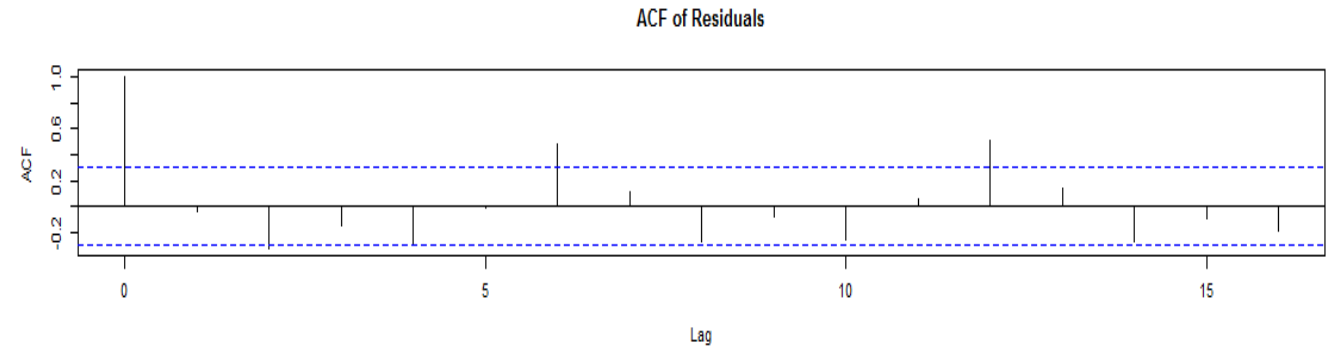
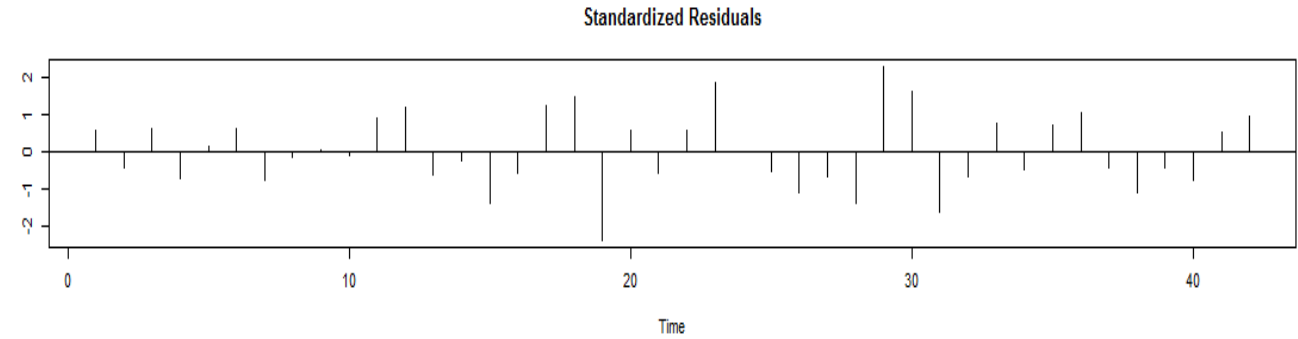
Series APAC_Consumer_quantity_local_pred



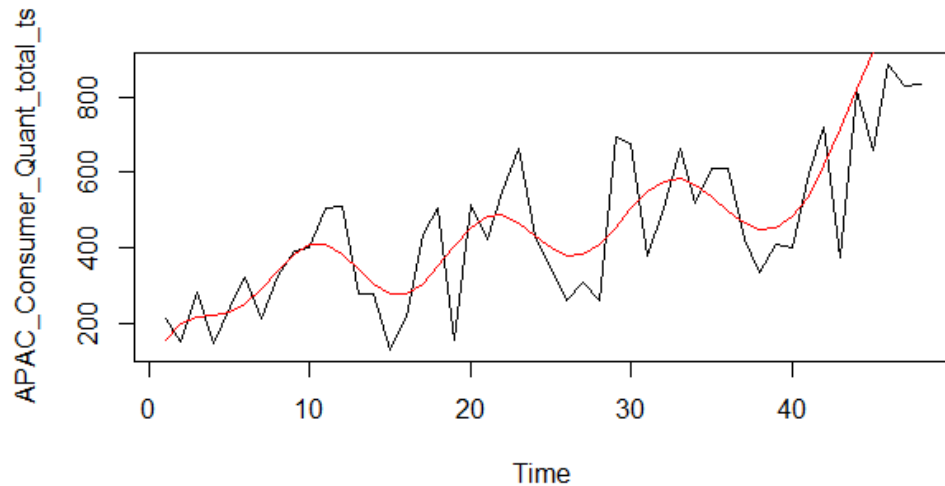
Model Diagnostics for a Fitted ARIMA Model - APAC – Consumer (quantity)

Series: APAC_Consumer_quantity_local_pred
ARIMA(0,0,0) with zero mean

sigma^2 estimated as 11108: log likelihood=-255.22
AIC=512.44 AICc=512.54 BIC=514.18



The Actuals Vs predictions

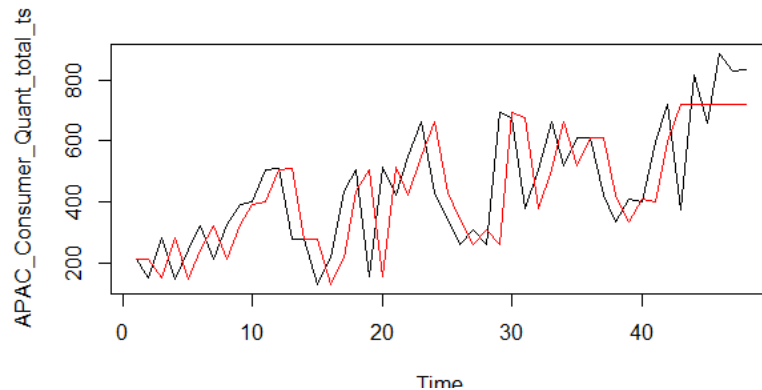


Model Diagnostics for a Fitted ARIMA Model :: APAC – Consumer (quantity)

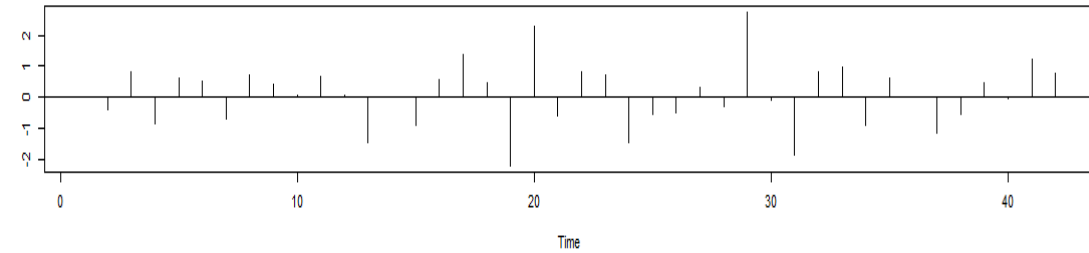
Series: APAC_Consumer_quantity_ts
ARIMA(0,1,0)

sigma² estimated as 25366: log likelihood=-266.07
AIC=534.14 AICC=534.24 BIC=535.85

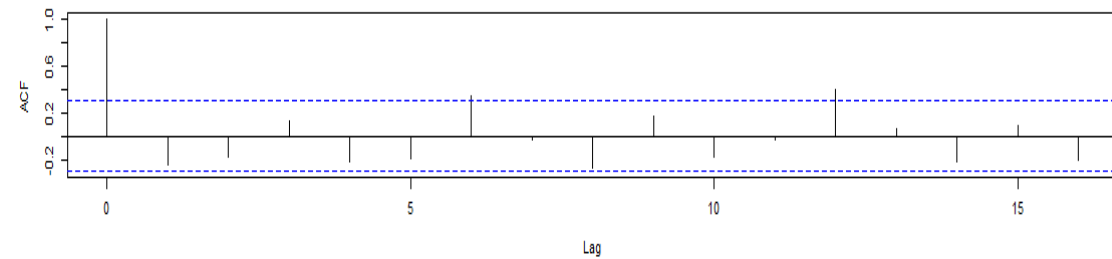
Actual Vs Auto ARIMA



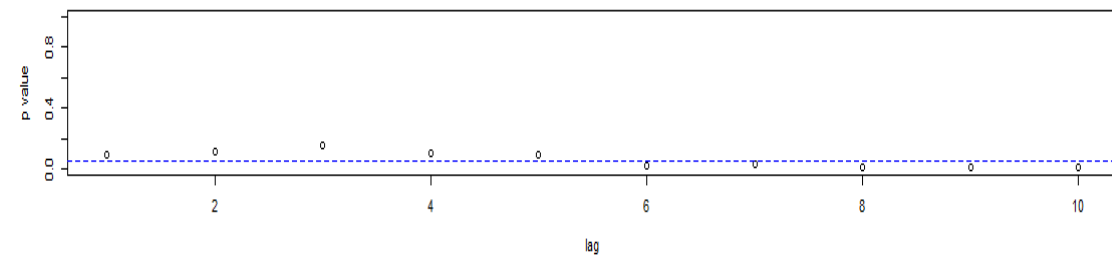
Standardized Residuals



ACF of Residuals



p values for Ljung-Box statistic



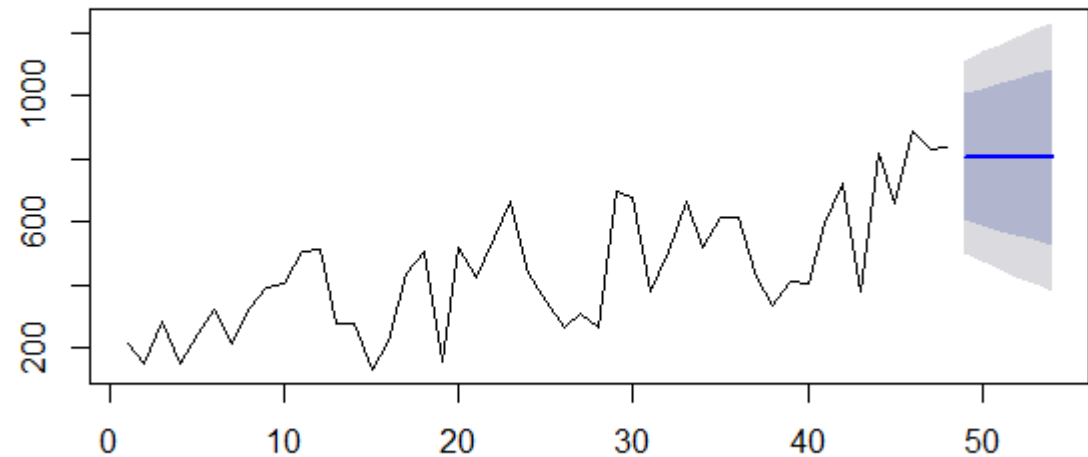
Conclusion:

The Auto ARIMA model is doing good in terms of MAPE value. it is lower than the classical decomposition method.

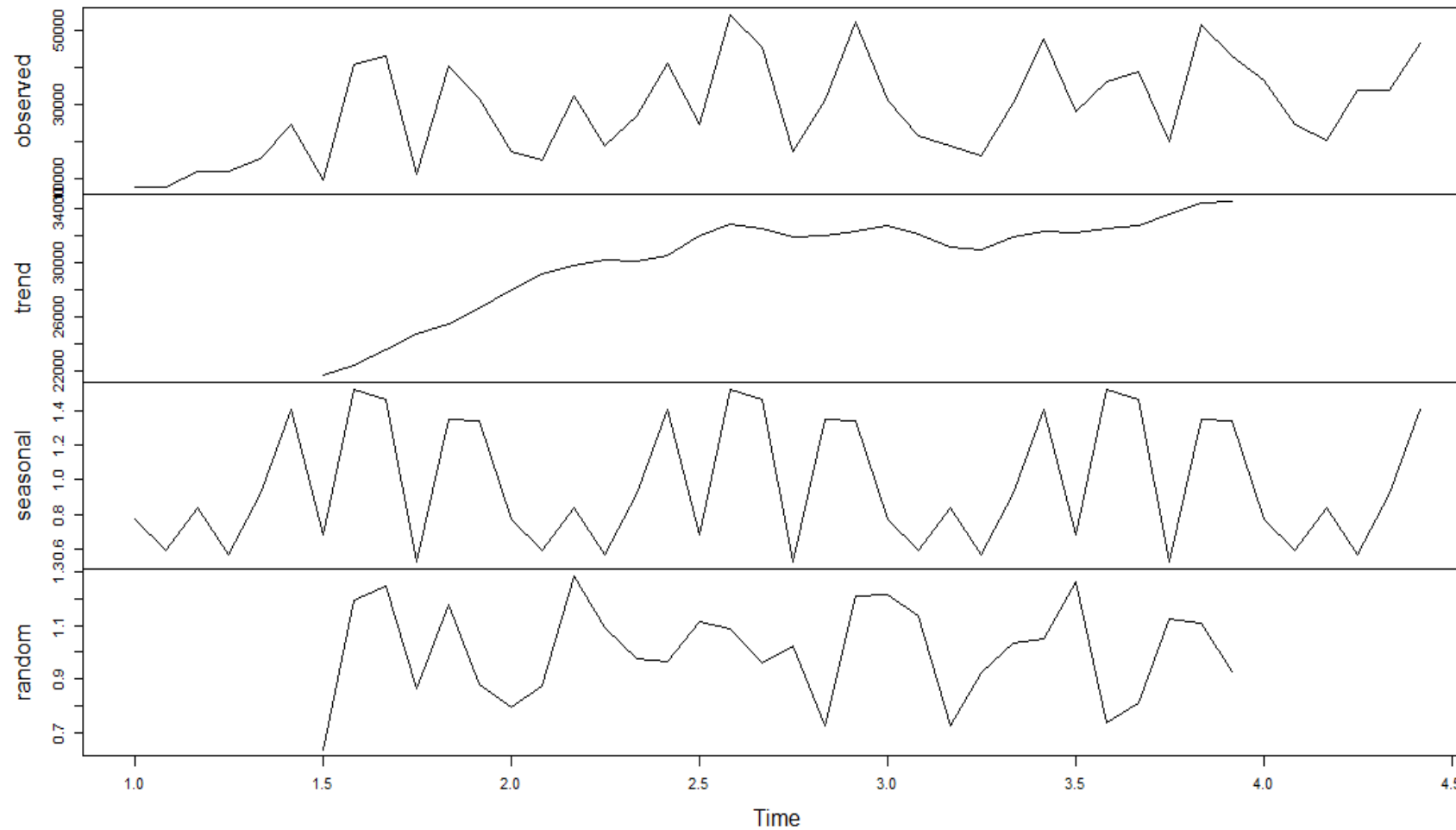
As ARIMA model is doing better than classical decomposition ,Forecast values for APAC Consumer Quantity for next 6 Months using ARIMA model

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
49	804.4143	604.3239	1004.505	498.4025	1110.426
50	804.4143	586.6355	1022.193	471.3504	1137.478
51	804.4143	570.2796	1038.549	446.3362	1162.492
52	804.4143	554.9940	1053.835	422.9588	1185.870
53	804.4143	540.5925	1068.236	400.9337	1207.895
54	804.4143	526.9375	1081.891	380.0501	1228.778

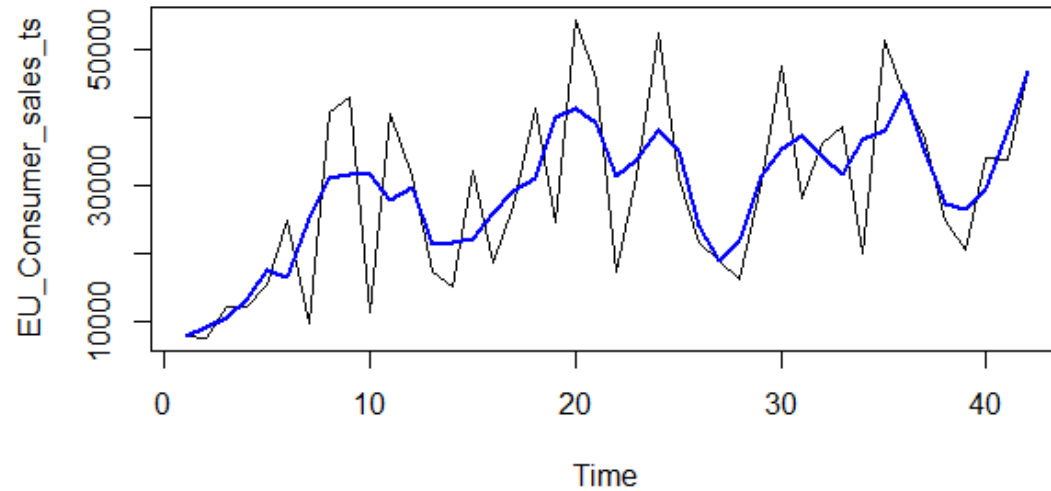
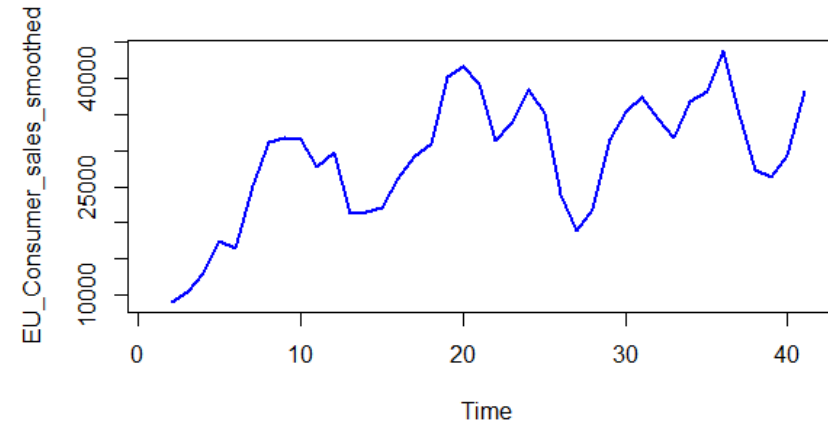
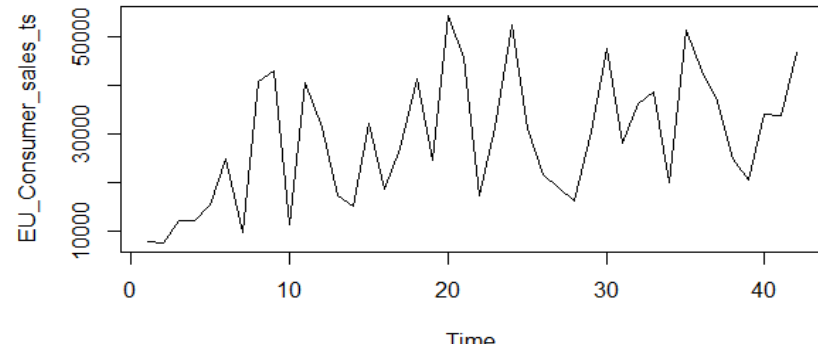
Forecasts from ARIMA(0,1,1)



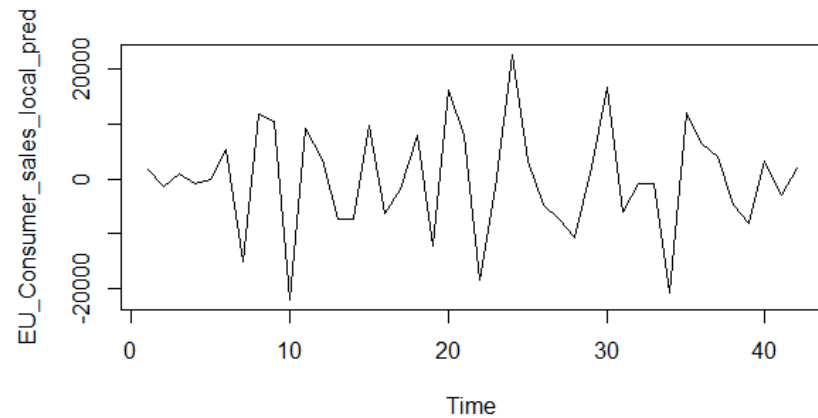
Decomposition of multiplicative time series



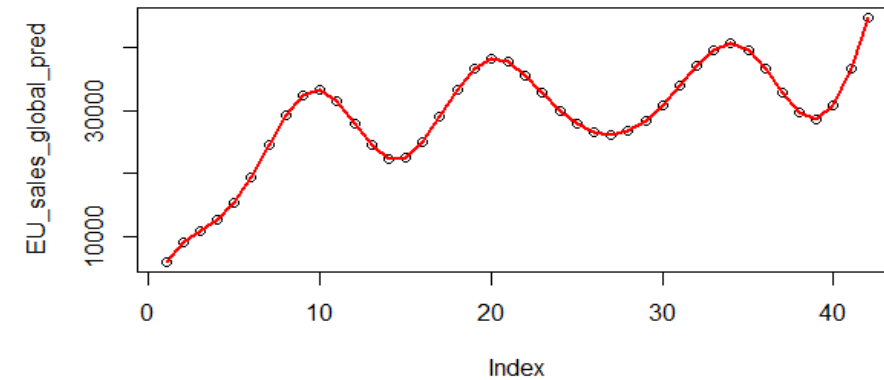
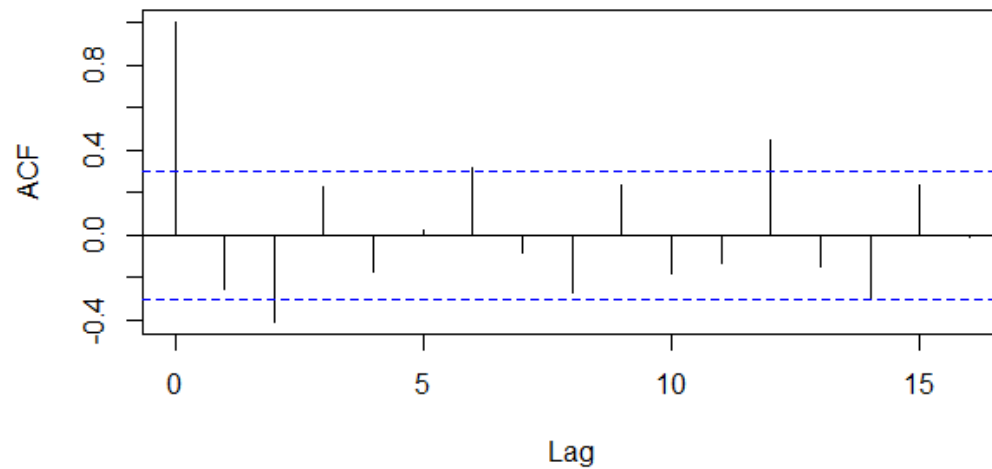
EU Consumer – Sales



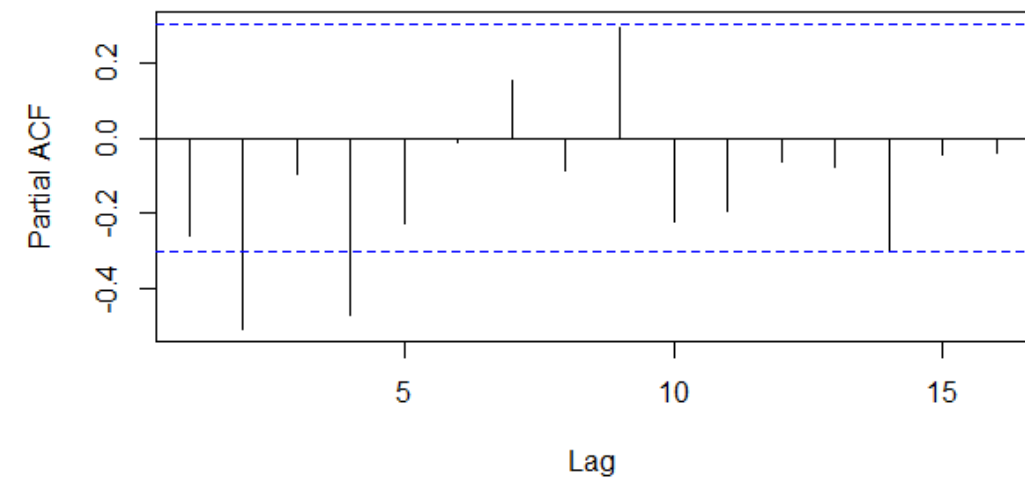
EU – Consumer (Sales)



Series EU_Consumer_sales_local_pred



Series EU_Consumer_sales_local_pred

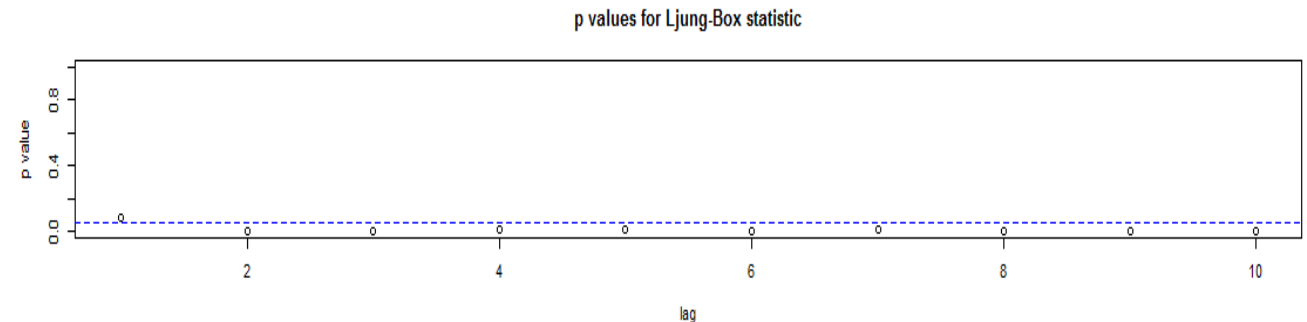
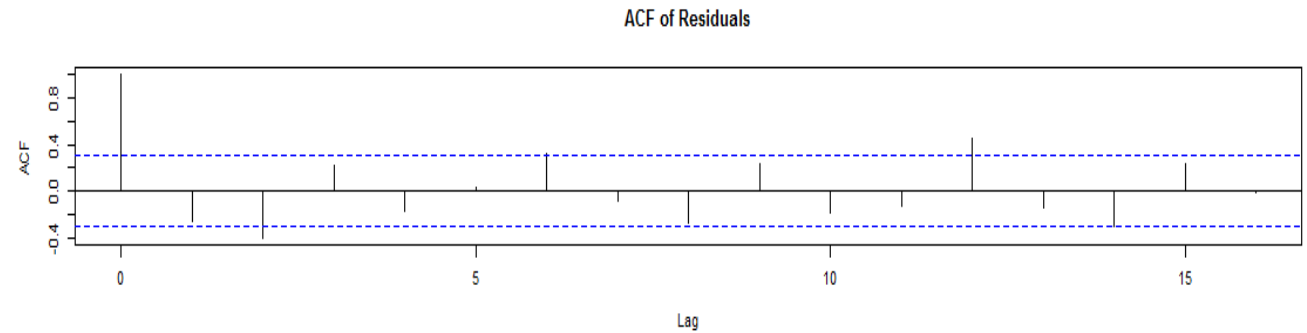
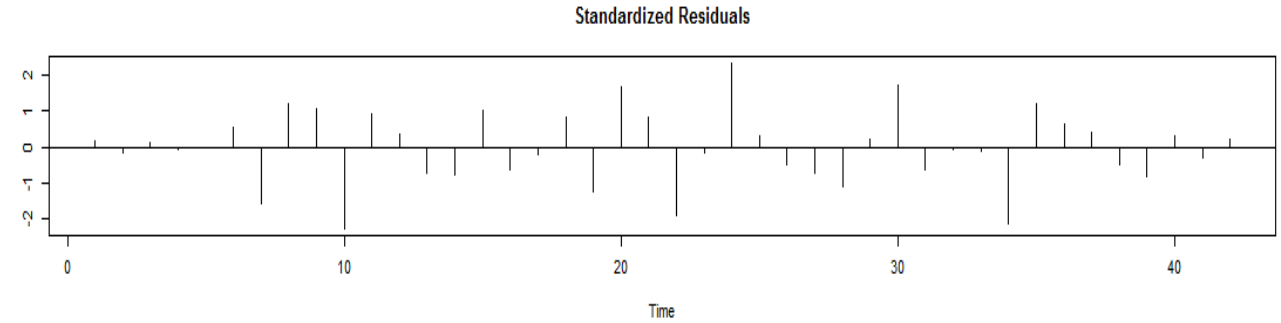
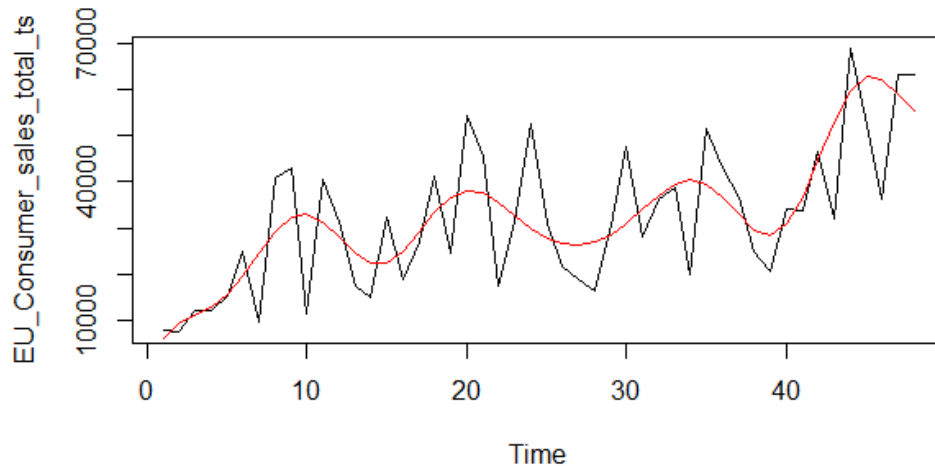


Model Diagnostics for a Fitted ARIMA Model :: EU – Consumer (Sales)

Series: EU_Consumer_sales_local_pred
ARIMA(0,0,0) with zero mean

sigma² estimated as 94396491: log likelihood=-445.22
AIC=892.44 AICC=892.54 BIC=894.18

The Actuals Vs predictions



Model Diagnostics for a Fitted ARIMA Model :: EU – Consumer (Sales)

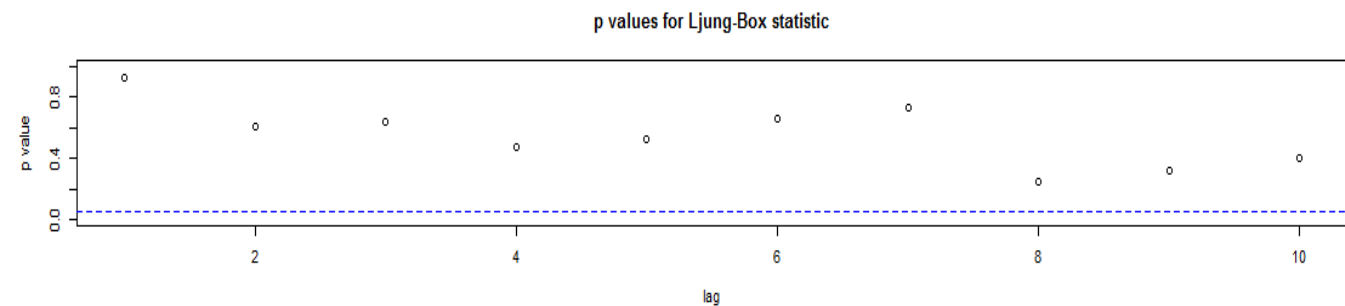
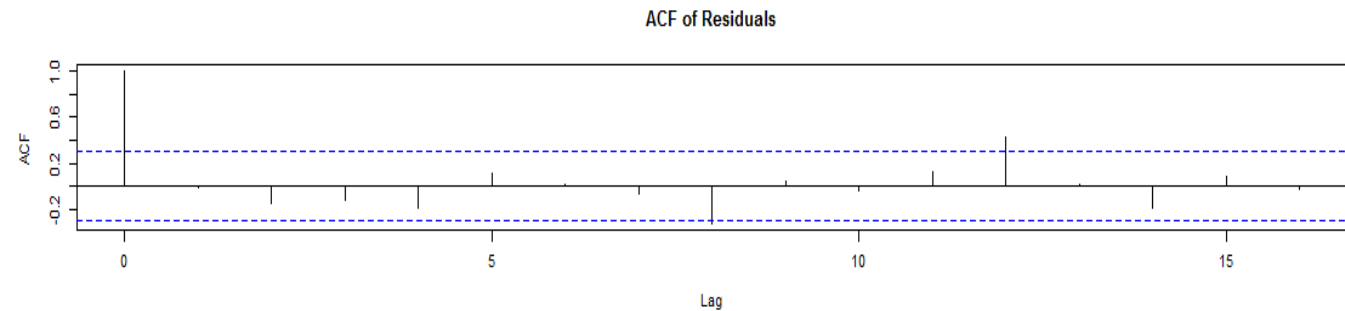
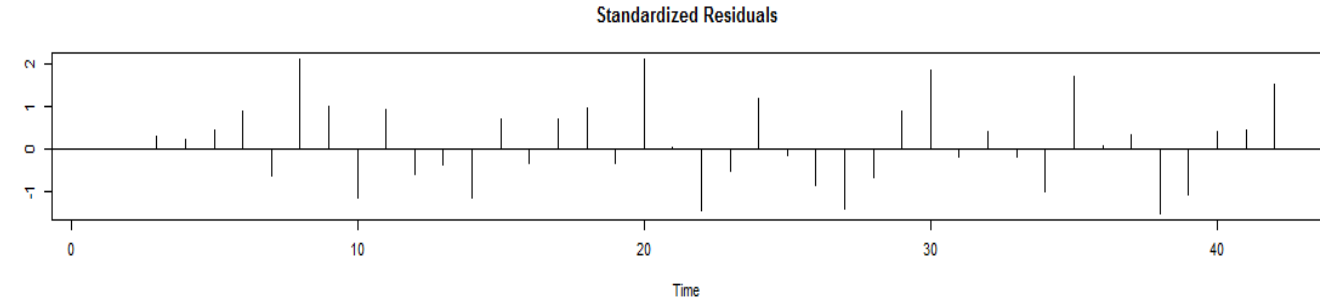
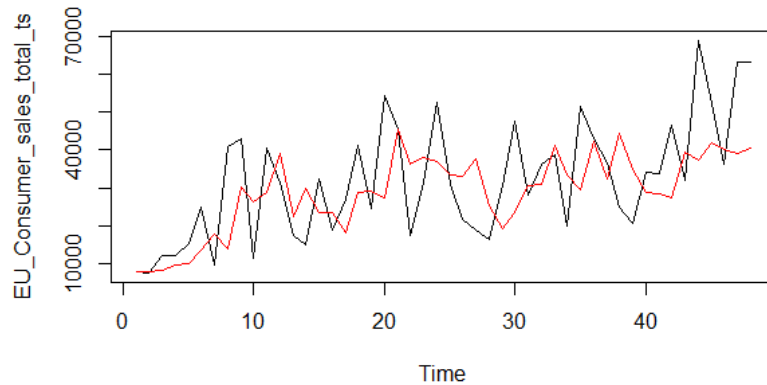
ARIMA(2,1,0)

Coefficients:

	ar1	ar2
coefficient	-0.5796	-0.4906
s.e.	0.1346	0.1310

sigma^2 estimated as 168564623: log likelihood=-445.84
AIC=897.67 AICC=898.32 BIC=902.81

Actual Vs Auto ARIMA

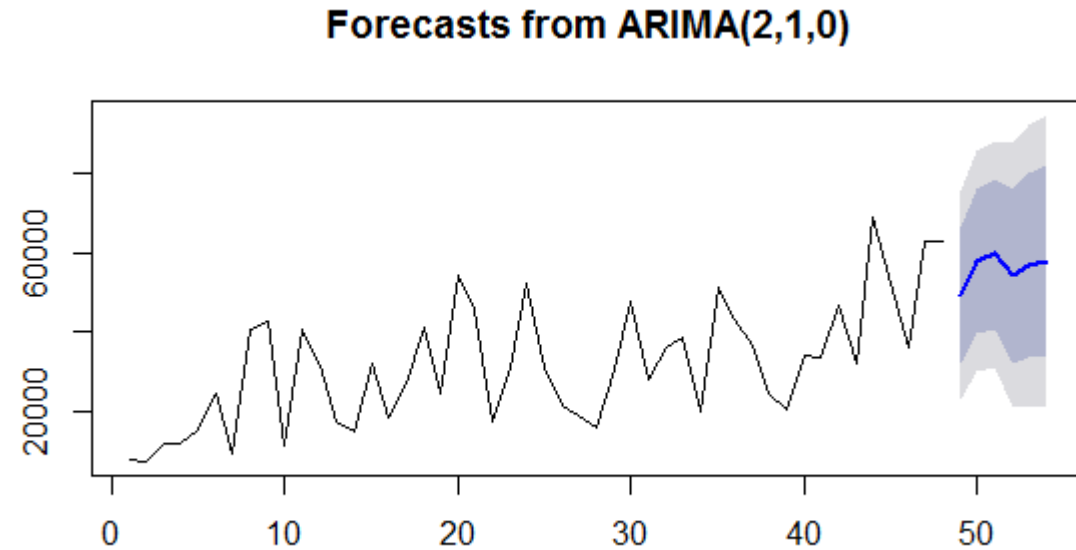


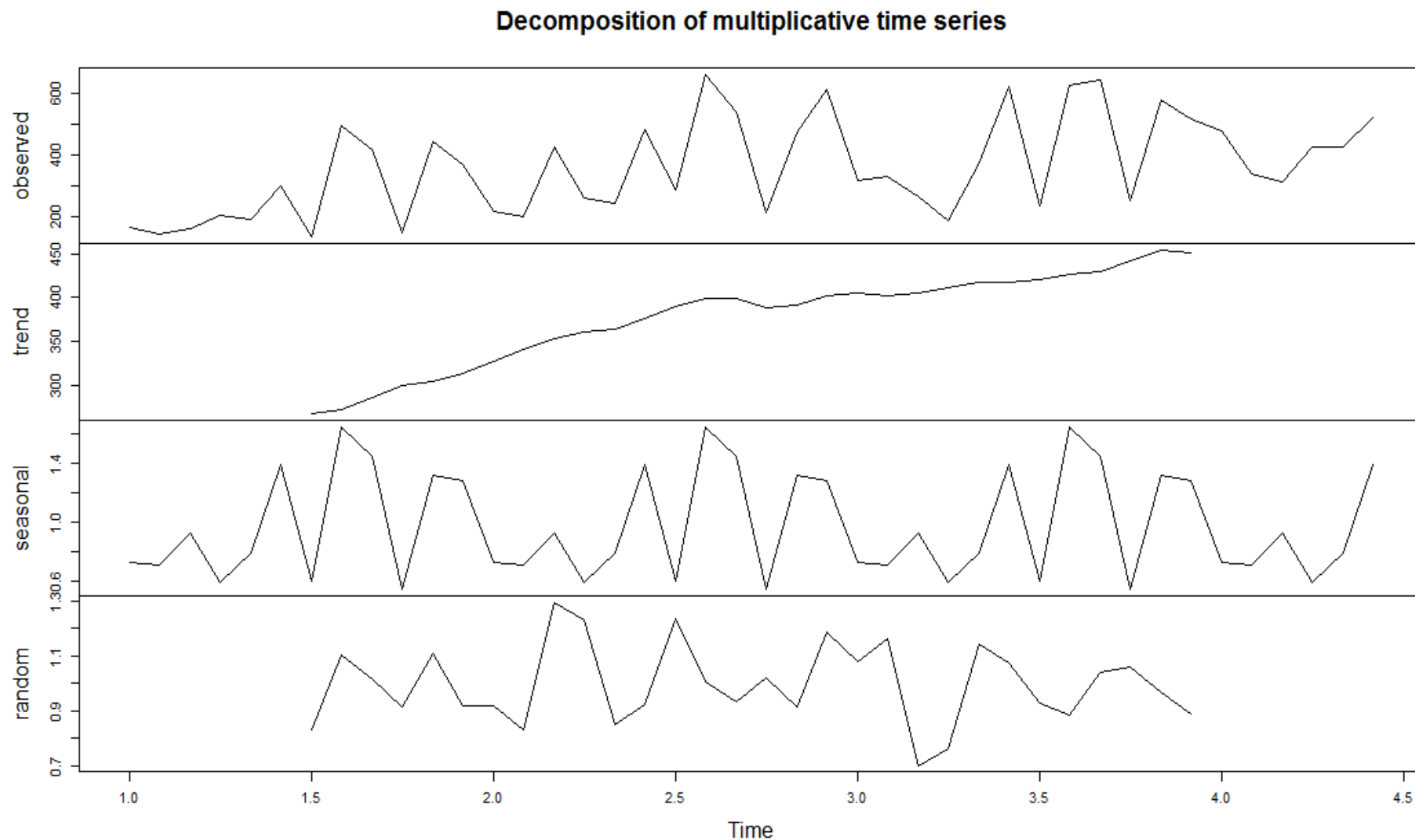
Conclusion:

The Auto ARIMA model is doing better in terms of MAPE value. it is lower than the classical decomposition method.

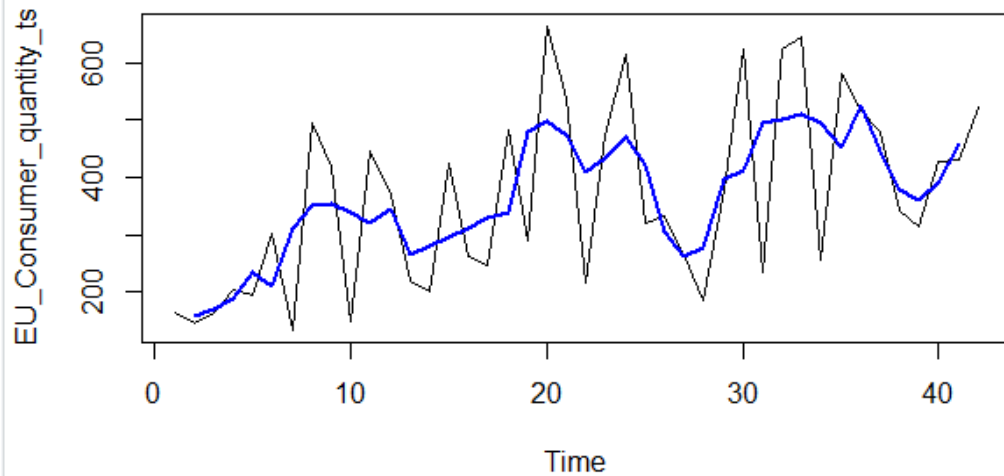
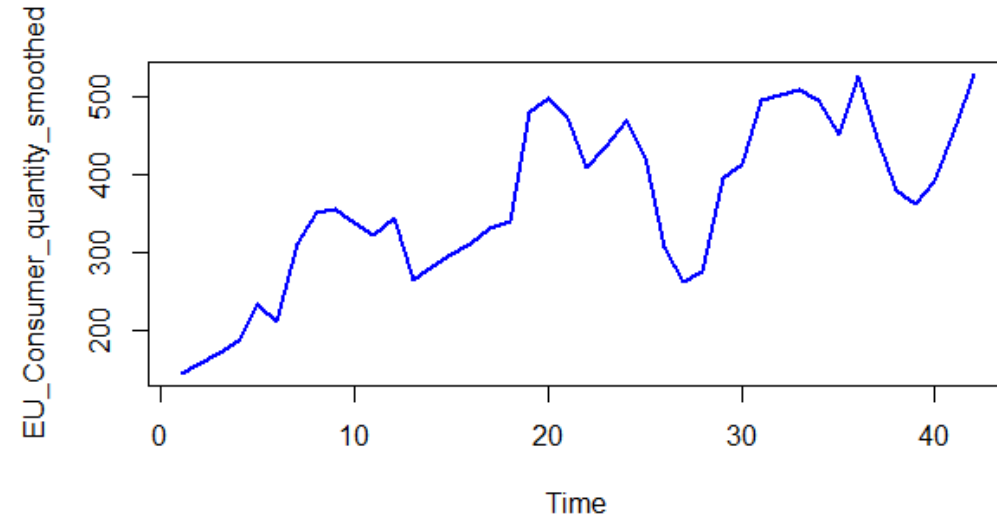
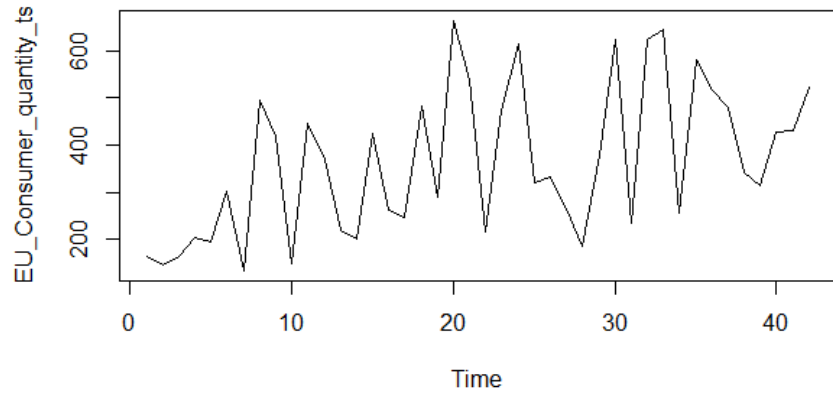
As ARIMA model is doing better than classical decomposition ,Forecast values for EU Consumer Sales for next 6 Months using ARIMA model

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
49	49358.71	32262.96	66454.45	23213.02	75504.39
50	58063.62	39825.62	76301.63	30171.00	85956.24
51	59714.33	40977.43	78451.22	31058.71	88369.94
52	54191.79	32401.83	75981.76	20866.93	87516.66
53	56811.55	33636.13	79986.98	21367.80	92255.31
54	58010.84	33966.75	82054.93	21238.57	94783.11

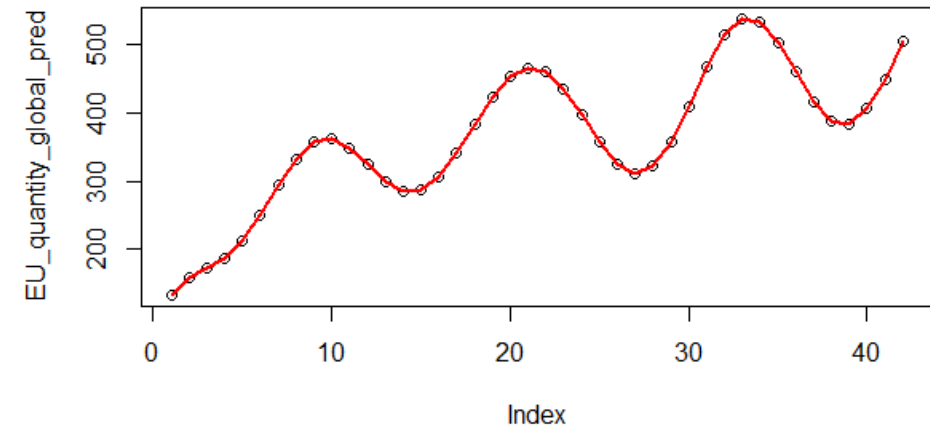
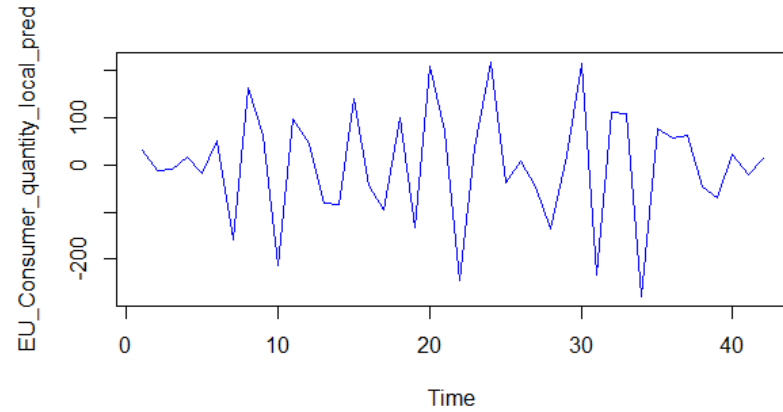




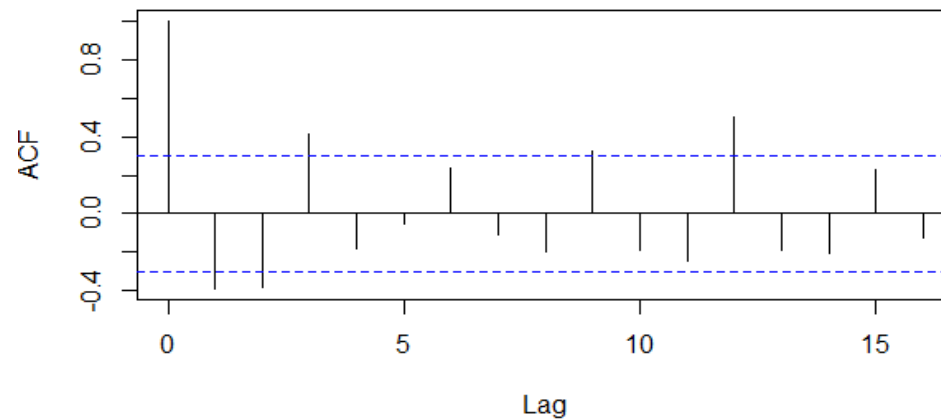
EU Consumer – Quantity



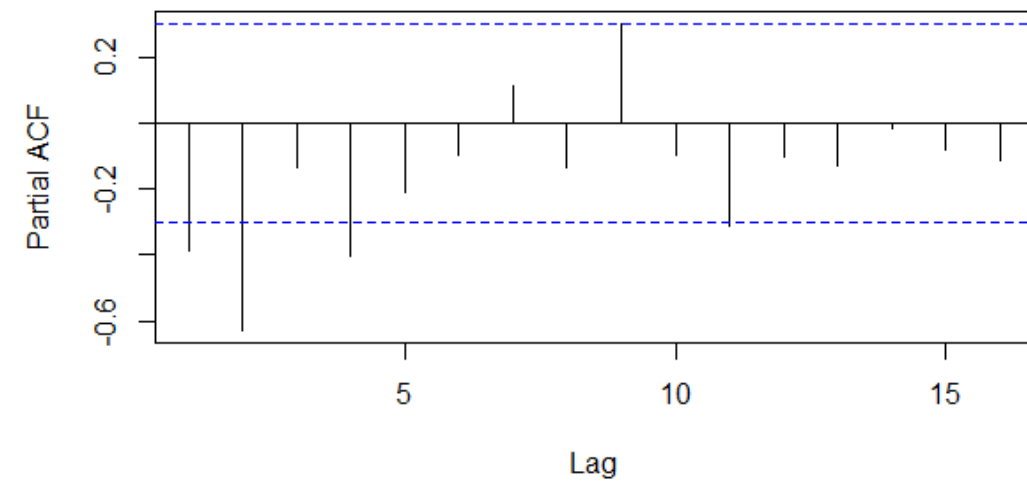
EU – Consumer (Quantity)



Series EU_Consumer_quantity_local_pred



Series EU_Consumer_quantity_local_pred



Model Diagnostics for a Fitted ARIMA Model :: EU – Consumer (Quantity)

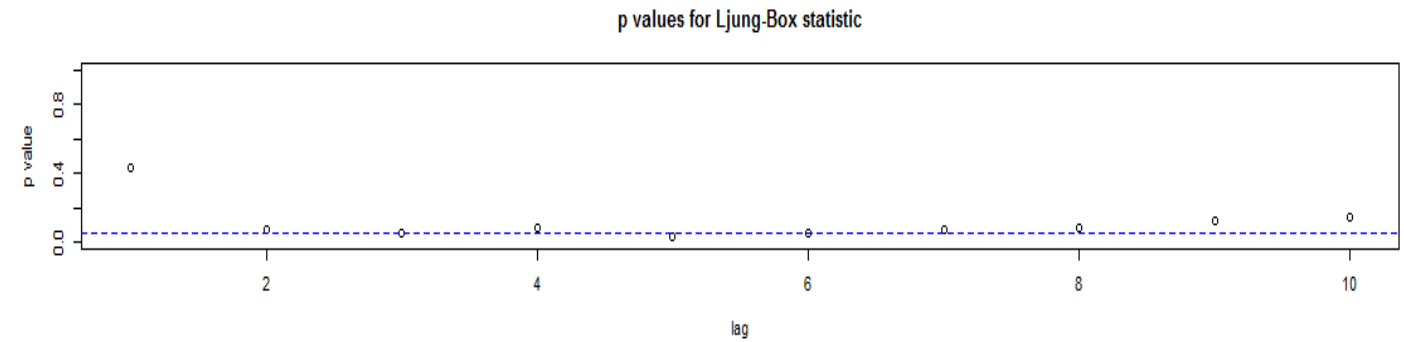
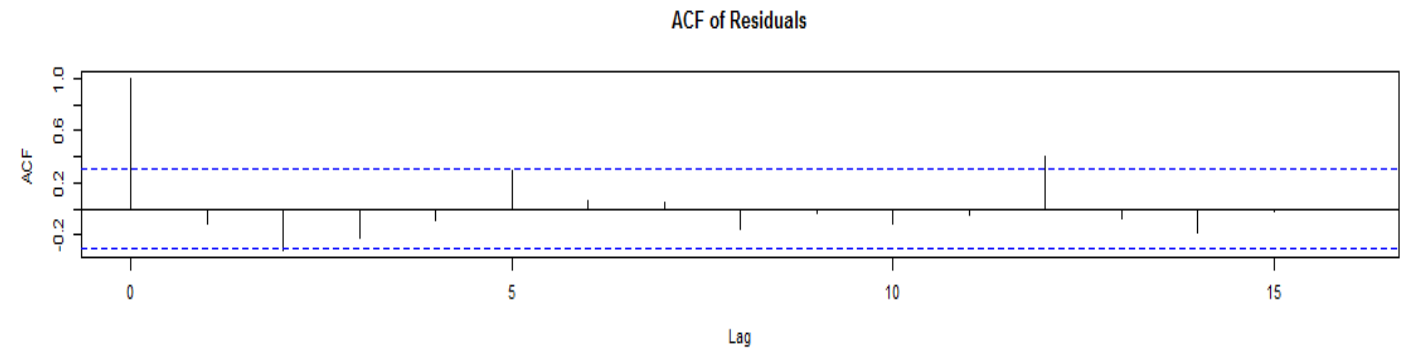
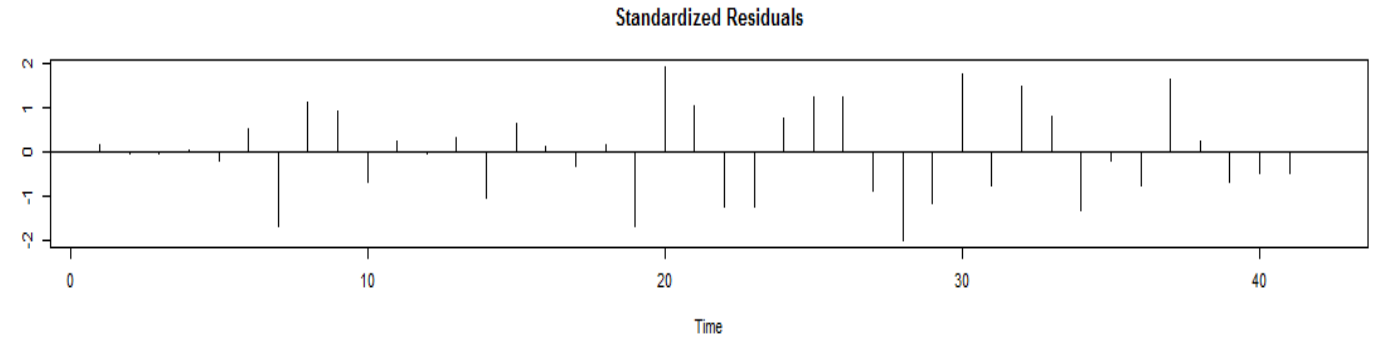
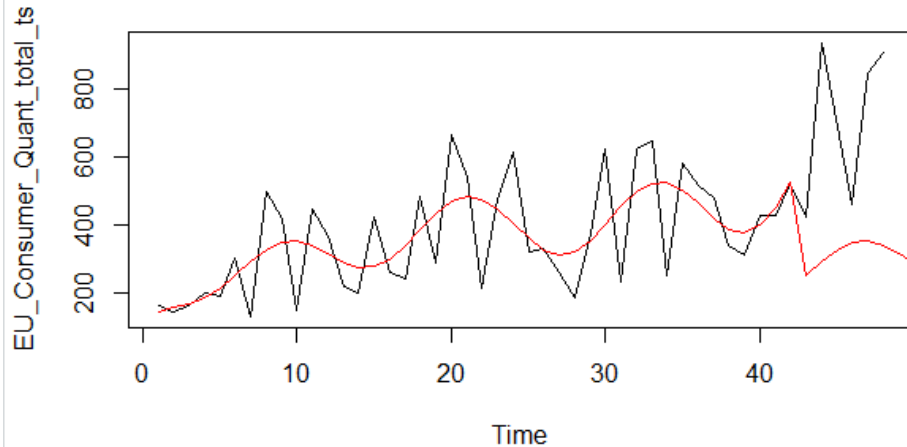
ARIMA(2,0,0) with zero mean

Coefficients:

	ar1	ar2
Coefficients:	-0.6341	-0.6158
s.e.	0.1173	0.1131

sigma² estimated as 7284: log likelihood=-245.89
AIC=497.79 AICC=498.42 BIC=503

The Actuals Vs predictions



Model Diagnostics for a Fitted ARIMA Model :: EU – Consumer (Quantity)

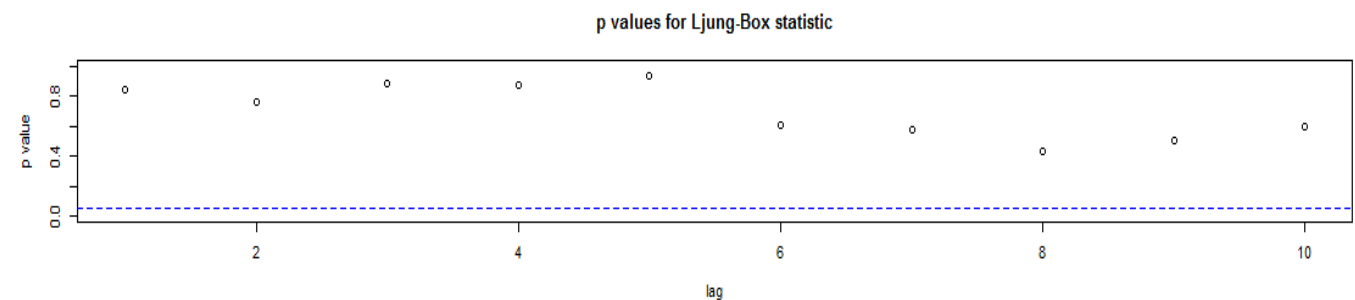
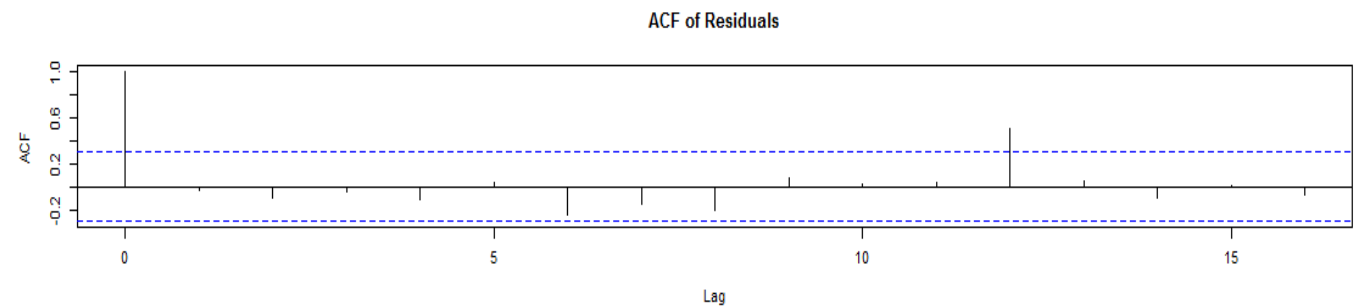
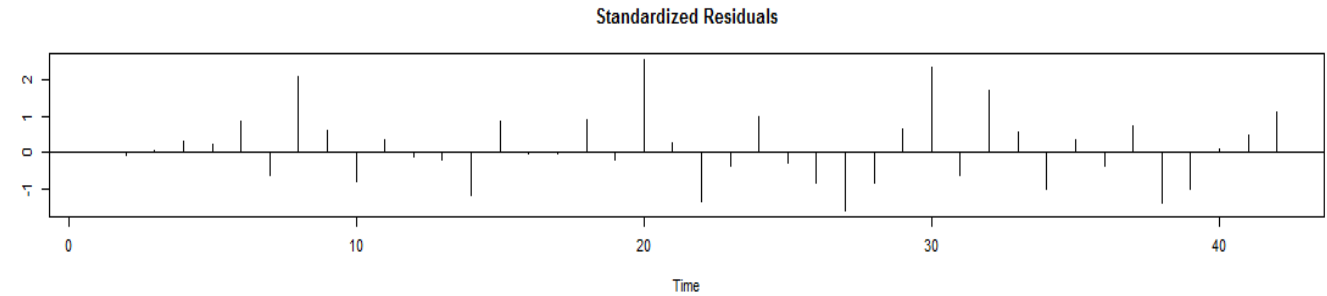
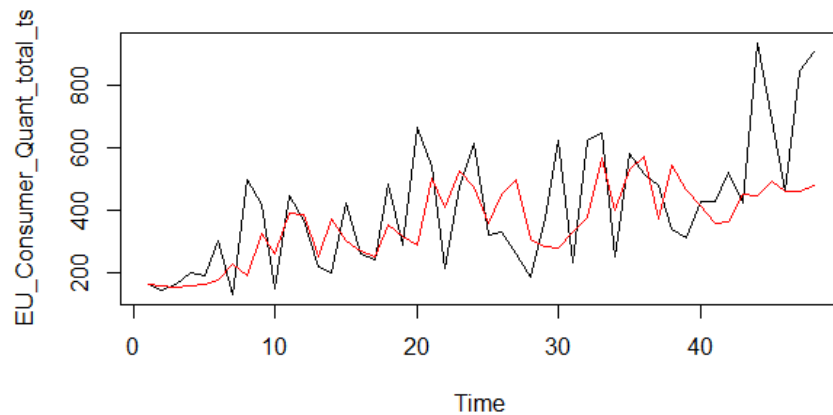
ARIMA(2,1,0)

Coefficients:

	ar1	ar2
coefficient	-0.7359	-0.5879
s.e.	0.1224	0.1185

sigma² estimated as 21185: log likelihood=-261.9
AIC=529.8 AICC=530.44 BIC=534.94

Actual Vs Auto ARIMA



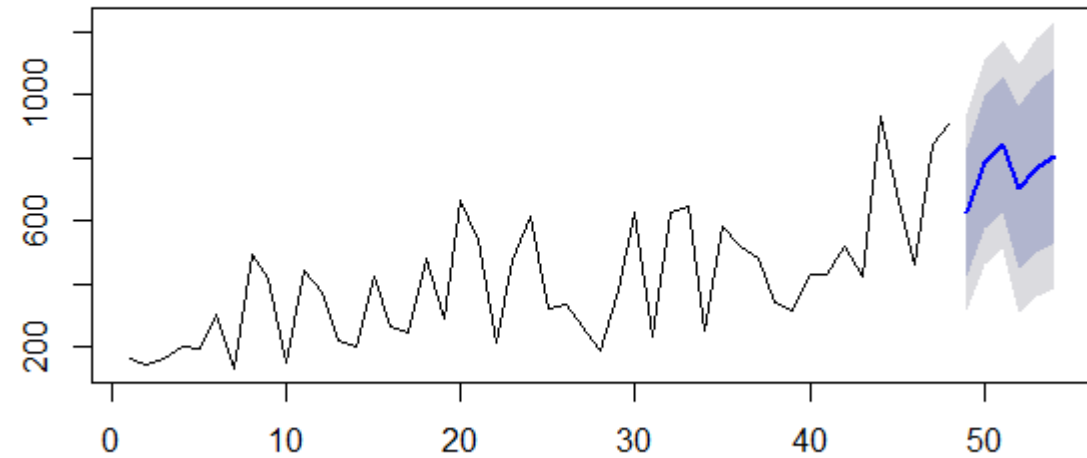
Conclusion:

The Auto ARIMA model is doing good in terms of MAPE value. it is lower than the classical decomposition method.

As ARIMA model is doing better than classical decomposition ,Forecast values for EU Consumer Quantity for next 6 Months using ARIMA model

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
49	626.2009	423.1689	829.2329	315.6903	936.7115
50	786.6056	575.2797	997.9316	463.4105	1109.8008
51	842.9179	628.3354	1057.5004	514.7423	1171.0935
52	704.8258	448.2217	961.4298	312.3838	1097.2678
53	768.6274	499.7864	1037.4683	357.4707	1179.7840
54	807.6497	532.1958	1083.1036	386.3793	1228.9201

Forecasts from ARIMA(2,1,0)



Based on historical given data set from 2011 – 2014 from Global Mart, we have find out 2 consistently profitable segments

- APAC – Consumer
- EU- Consumer

Build and Evaluated model for above segments for Sales and Quality.

- APAC-Consumer-Sales :- classical decomposition is doing better then ARIMA based on MAPE values.
- APAC-Consumer-Quality :- ARIMA is doing the best in terms of MAPE values
- EU-Consumer-Sales :- ARIMA is best model.
- EU- Consumer- Quality :- ARIMA is doing good then classical decomposition.

Forecasted next 6 months (1/2015 to 6 /2015) based on the best model.

Forecast observations for the next 6 Months :

1. APAC-Consumer-Sales : There is no much raising of the sales for next 6 months
2. APAC-Consumer-Quality : its constant raise in the next 6 months
3. EU-Consumer-Sales : slight raining the sales data.
4. EU-Consumer-Quality : initially it started with low value and then increased and there is a slight drop in April -2015 and again it reached the target.